FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Towards the Identification of Psychophysiological States in EEG

Helena Margarida de Gouveia Faria

DISSERTATION FOR THE DEGREE OF MASTER IN BIOENGINEERING

MSC IN BIOENGINEERING - BIOMEDICAL ENGINEERING

Supervisor: João Paulo Trigueiros da Silva Cunha, PhD.

6th July 2018

Towards the Identification of Psychophysiological States in EEG

Helena Margarida de Gouveia Faria

MSc in Bioengineering - Biomedical Engineering

6th July 2018

Resumo

Os processos fisiológicos e cognitivos no corpo humano são regulados por mecanismos que se comportam de forma indissociável, apresentando uma clara inter-dependência. Se por um lado, a fisiologia é muitas vezes um reflexo dos processos cognitivos. Por outro, mudanças nos mecanismos fisiológicos provocam alterações nos pensamentos, comportamentos e emoções. Catapultado pelo desenvolvimento de tecnologias como a realidade virtual e os dispositivos *wearable* para monitorização de variáveis fisiológicas, o estudo das respostas fisiológicas a estados emotivos tem um vasto leque de possíveis aplicações, em áreas tão diversas como a medicina e *eSaúde*, a educação, o entretenimento ou o marketing.

Esta dissertação teve como um dos principais objetivos o desenvolvimento de um sistema capaz de reconhecer quatro emoções em sinais EEG: alegria, calma, tristeza e medo. Para tal, foi seguida uma abordagem típica de classificação: após um pré-processamento dos sinais, foram extraídas diversas características, com base nos métodos mais comuns na literatura. As características mais relevantes para a discriminação das diferentes emoções foram selecionadas recorrendo ao algoritmo ReliefF e serviram como entrada de um classificador SVM com kernel do tipo RBF. Uma base de dados disponível publicamente (DEAP) foi usada como forma de validar os resultados. Apesar de uma comparação direta com o estado da arte ser difícil devido à diversidade de abordagens seguidas, a exatidão obtida de 78,71% para reconhecimento de 4 emoções revela um desempenho superior a alguns dos trabalhos apresentados na literatura. A abordagem seguida mostrou-se robusta, sendo eficaz em novos dados, adquiridos em diferentes condições.

A segunda parte deste trabalho focou-se na influência do tipo de estímulo visual na perceção e expressão das mesmas emoções. Neste sentido, foi desenhado um protocolo de estimulação emocional com três modalidades de estímulos: imagens, vídeos 2D e vídeos esféricos num ambiente de realidade virtual. Para cada caso, os participantes avaliaram a sua experiência de acordo com a escala SAM em termos de *valência* e *excitação*. 17 voluntários participaram neste estudo, sendo que foram recolhidos os sinais EEG de 8 dos participantes. No geral, apesar de se terem encontrado algumas diferenças pontuais entre modalidades, a análise aos resultados das avaliações dos participantes não revelou um efeito predominante de nenhum dos tipos de estímulo. No entanto, numa análise exploratória aos valores de potência para as bandas Alfa, Beta e Teta dos sinais EEG dos participantes, verificou-se que os vídeos em realidade virtual tiveram um grande impacto na sua atividade cerebral. Este facto foi particularmente notório para as emoções com valores elevados de excitação, indicando que essa dimensão emocional, apesar de mais difícil de compreender e expressar, parece ser bem captada no sinal EEG.

Apesar das dificuldades inerentes a um tópico tão subjectivo como são as emoções, o trabalho desenvolvido nesta dissertação corrobora a utilização de sinais fisiológicos para a identificação de estados emocionais, com resultados encorajadores. Estes apontam ainda para os benefícios que podem ser retirados da utilização de sistemas de realidade virtual em contextos onde a emoção tenha um papel relevante e indicam uma interessante direção de investigação para o futuro.

Abstract

The mechanisms that regulate our physiological and mental processes behave in a coupled way in which there is a notorious interdependency. Mental processes are responsible for changes in the physiological state of our body. On the other side, changes in bodily functions also lead to different thoughts, behaviours and emotions. Boosted by the development of technologies such as virtual reality and wearable devices, understanding the physiological responses to emotional states can serve a wide range of valuable applications in such diverse fields as medicine and *eHealth*, education, entertainment or marketing.

One of the main goals of this dissertation was to develop a system for emotion recognition in EEG signals. A typical classification pipeline was followed: after a preprocessing stage, a range of features were extracted from the signals. The most important features for discriminating the four target emotions (joy, calmness, sadness and fear) were selected using the ReliefF algorithm. These were fed to a SVM classifier with a RBF kernel. The proposed methodology was tested in a publicly available dataset (DEAP). Although a direct comparison with the state-of-the-art is difficult due to myriad of approaches followed by different authors, the high accuracy of 78,71% obtained in this dissertation reveals a better performance than some of the works in the literature. Furthermore, the same methodology proved to be robust and efficient in new data, recorded in different conditions.

The second part of this work was focused on the influence of the modality of visual stimuli in the perception and expression of the same set of emotions. Therefore, a stimulation protocol was designed so as to elicit specific emotional responses with images, 2D videos and spherical videos in a virtual reality environment. The participants rated their experience using the SAM scale, representing *valence* and *arousal*. 17 volunteers took part in this study. The EEG was recorded during the experiment for 8 of them. When analysing the ratings given by the participants, only some one-off differences were found between modalities. Overall, the results did not reveal any predominant effect from any of the modalities. However, the exploratory analysis conducted to investigate the power of the Alpha, Beta and Theta frequency bands indicated that the virtual reality videos had a significant impact on the participants' brain activity. This was particularly noticeable for the high-arousal emotions, which shows that the effects of this emotional dimension, despite being harder to recognize and express, can be captured in the EEG signal. That is, EEG features can capture emotional responses which can not be expressed in the form of a rating.

In spite of the obstacles in such an intrinsically subjective topic as are emotions, the work developed in this dissertation validates the use of physiological signals to identify emotional states with encouraging results. These results also point to the benefits of using virtual reality in contexts where emotion is of paramount importance and indicate a line of research which will surely bear fruit in real applications.

Agradecimentos

As minhas primeiras palavras de agradecimento dirigem-se ao Professor João Paulo Cunha, principalmente pela sua diponibilidade, orientação e ideias durante a realização desta tese. O obrigado estende-se a todos os elementos do BRAIN pelo ambiente amigável em que me incluíram e que permitiu tornar mais leve este semestre. Agradeço também a todos os participantes envolvidos no estudo, pois sem eles a realização deste trabalho não seria possível .

Chegado ao fim o meu percurso académico, agradeço a todos os professores que durante 17 anos me moldaram enquanto aluna, mas principalmente enquanto pessoa. À minha professora primária, por me ter incutido, desde cedo, que *podemos sempre ser melhores*.

Obviamente não posso deixar de me dirigir à minha família, por tudo o que depositaram em mim, desde sempre. Nunca as palavras poderão traduzir o meu agradecimento.

Por último, impõe-se uma enorme palavra de apreço àqueles que me acolheram, me ensinaram e inspiraram ao longo do meu percurso na FEUP. A todos os que fizeram destes 5 anos *um tempo que não passa, neste passar do tempo que não volta*. A um grupo que me mostrou valores, momentos e pessoas, numa casa que não era a minha. Aos que partilharam comigo os mesmos paralelos, aos que defenderam as mesmas cores. Aos de sempre, aos de ontem e aos de amanhã. Por todas as noites frias e as tardes de calor. Pelas lágrimas e pelas gargalhadas. Pelas conversas e pelas histórias. Pelo Fado e pelo *trashy*. Pela tradição e pela inovação. Pelo bom e pelo mau. Pelo côncavo e o convexo. Pelo branco e pelo preto. A Metal&Bio.

Obrigada.

Helena Margarida

"Hear this, young men and women everywhere, and proclaim it far and wide. The earth is yours and the fullness thereof. Be kind, but be fierce. You are needed now more than ever before. Take up the mantle of change. For this is your time."

Sir Winston Churchill

Contents

List of Figures

List of Tables

Abbreviations

ANS	Autonomic Nervous System
BCI	Brain-Computer Interface
ECG	Electrocardiogram
EDA	Electrodermal activity
EEG	Electroencephalogram
EMG	Electromyogram
HCI	Human-Computer Interaction
HRV	Heart Rate Variability
HVHA	High-Valence, High-Arousal
HVLA	High-Valence, Low-Arousal
IAPS	International Affective Picture System
k-NN	k-Nearest Neighbours
LVHA	Low-Valence, High-Arousal
LVLA	Low-Valence, Low-Arousal
PNS	Peripheral Nervous System
PTSD	Post Traumatic Stress Disorder
SAM	Self-Assessment Manikin
SVM	Support Vector Machine
SCWT	Stroop Colour Word Test
TSST	Trier Social Stress Test
VR	Virtual Reality

Chapter 1

Introduction

1.1 Context

Mens sana in corpore sano. This old latin saying can be thought of as the most basic premise of psychophysiology. The mechanisms that regulate our physiological and mental processes behave in a coupled way in which there is an indisputable dependency. Our mental processes are responsible for alterations in the physiological state of our body, while changes in physiology lead to different thoughts, feelings and behaviour (Critchley, 2009). One of the clearest examples of this mutual interaction is the variety of physiological responses triggered by one's emotions. This is a major aspect of human personality, since emotions are involved in decision-making, interaction and intelligence (Atkinson and Campos, 2016).

Defining emotion has long been a controversial theme amongst academics (Scherer, 2005; Thompson, 1994). However, the concept is deeply grounded in everyday use and can be defined as a psychophysiological process triggered by conscious and/or unconscious perception of an object or situation. It is commonly associated with mood, temperament, personality, disposition and motivation (Koelstra et al., 2012). There are several theories and models for classifying emotions, yet two of them have gained special interest – the discrete and the dimensional models. The discrete model is comprised of six basic emotions (happiness, sadness, fear, surprise, disgust and anger) from which all other derive or are part of. These are considered to be universally accepted (Ekman et al., 1987). The two-dimensional model proposed by Lang (1995) places emotions on two axis – valence and arousal, as shown in Figure 1.1. Valence defines the polarity of emotions, which can be positive or negative. Arousal denotes the intensity (high or low).

Emotions such as anger, anxiety or depression are frequently associated with stress. In fact, psychological stress is frequently considered as a type of emotion (Kim et al., 2004; Lazarus, 1993). The term stress in this context was first defined by Selye (1956) as the non-specific response of the body to any demand made upon it.

In low to moderate levels, stress can improve performance in an attempt to promote adaptation. However, continuous exposure to acute stress inflicts damage to physical and mental well-being. There is an increasing recognition that an excessive or long-lasting stress response to a

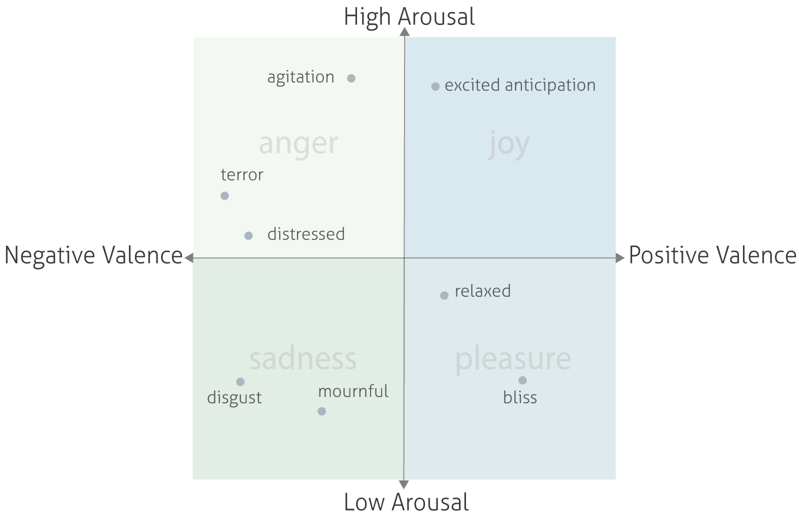


Figure 1.1: Valence-arousal plot of emotions. Adapted from Selvaraj et al. (2013)

variety of external or internal stressors is associated with the most common multi-factorial, non-communicable diseases affecting society (Nicolaides et al., 2015). Continued stress has been associated with anomalies in the immune system (Khansari et al., 1990) and may cause or worsen chronic cardiovascular diseases such as hypertension (Al'Absi and Arnett, 2000). Stress is also a risk factor for several neurological disorders and adverse behaviours such as depression, rage, anxiety, and addiction (McEwen and Stellar, 1993; Chrousos and Gold, 1992). It has also been more recently linked to Alzheimer's and Parkinson's disease (Rothman and Mattson, 2010; Djamshidian and Lees, 2014).

The economic impact of stress is also noteworthy. Exposure to stress was considered to be the main workplace health and safety risk, indicated by 53% of European workers (Milczarek, 2014). The total annual costs for the EU-15 countries[1] of work-related stress were estimated to be €20 billion (Hassard et al., 2014).

1.2 Motivation

Studying the physiological responses to psychological states such as emotions or stress is a challenging problem but has a great number of valuable applications. Agrafioti et al. (2012) claims that the design of systems which can perform automatic identification of human psychophysiological states, such as different emotions, would transform such diverse fields as medicine, entertainment, education or safety. A system that could detect and adapt itself according to the user's emotional state could be used for monitoring patients with mental health disorders and adjust their treatment. It would also bring benefits in applications where attention and motivation are crucial factors to be monitored. These include educational scenarios where the interest and engagement are fundamental for effective learning (Picard et al., 2004).

[1]EU-15: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom

Modelling emotions has also recently drawn a lot of interest in Human-Computer Interaction (HCI), more specifically in Brain-Computer Interfaces (BCI). In fact, extending the expression and understanding of emotions to machines and computers would provide sympathy, helping humans in the interaction with those machines, in a process commonly known as affective computing (Selvaraj et al., 2013; Al-Nafjan et al., 2017). According to the Gartner's 2017 Hype Cycle report on trending research topics, BCI is at the Innovation Trigger stage (Gartner, 2017), as shown in Figure 1.2. To date, several assistive applications have been developed, such as brain-controlled wheelchairs, keyboards or prosthetic devices(Galán et al., 2008; Millán et al., 2010). However, mainstream adoption is predicted to occur in more than 10 years, indicating that there is still a rather long way to pave (Gartner, 2017).

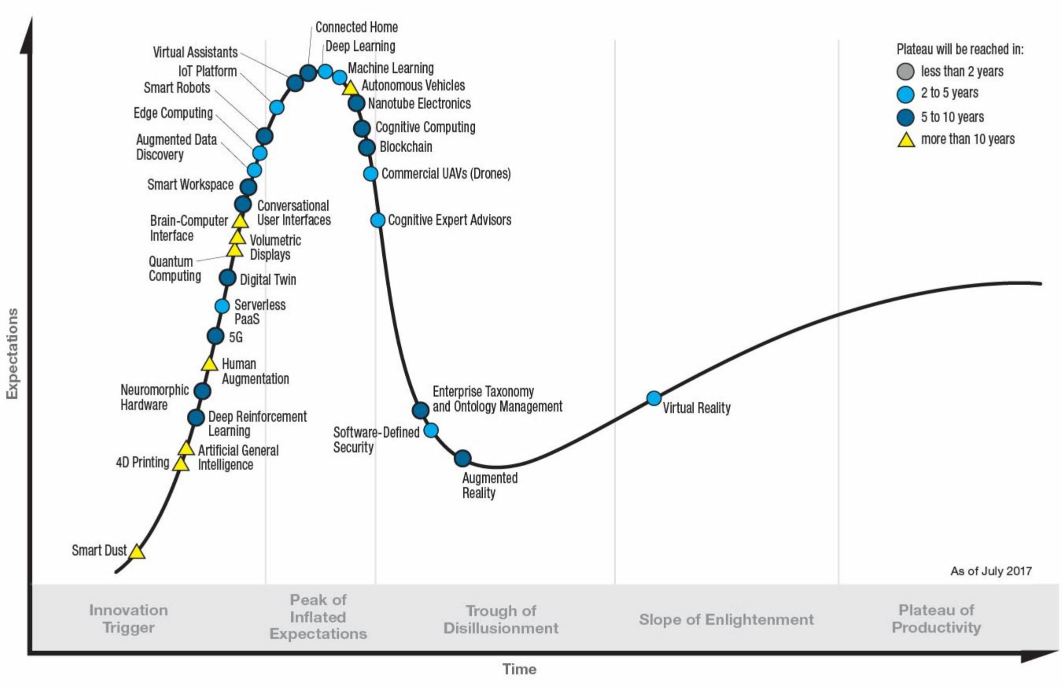


Figure 1.2: Gartner's Hype Cycle for Emerging Technologies 2017. From Gartner (2017)
.

Besides emotions in general, ubiquitous stress detection would enable individuals to manage and become more aware of their stress levels. In fact, real-time and continuous monitoring of stress in daily life can enhance stress awareness and provide valuable insights for further research on the topic. It has the potential to provide prompt intervention not only for stress management, but also to deal with other conditions induced by stress such as smoking, drinking, drug use or depression (Rahman et al., 2014). Stress monitoring is also relevant in military contexts where significant resources are spent, both in improving mental resiliency (Calibo et al., 2013) and in managing and coping with PTSD (Wood et al., 2006; Gaggioli et al., 2014).

Research on human psychophysiology has witnessed the use of several different physiological signals in order to determine better ways to measure and monitor stress. These biosignals include changes in heart rate, blood pressure, electrodermal activity and pupil diameter, as well as electroencephalography (Sharma and Gedeon, 2012; Calibo et al., 2013). However, the links between brain activity and stress in non-clinical context have not yet been fully understood (Hosseini and Khalilzadeh, 2010; Calibo et al., 2013). The recent surge of convenient devices capable of collecting physiological data has increased the popularity and interest in studying these signals and clearly shows a direction for further research (Pomer-Escher et al., 2014; Jun and Smitha, 2016).

1.3 Objectives

The main goal of this dissertation is to create a system for psychophysiological state assessment, more specifically intended for emotion recognition using brain activity. The potential of virtual reality for emotion induction in EEG studies will also be assessed by studying the response to visual stimuli in different modalities: images, 2D videos and spherical videos in virtual reality.

1.4 Structure

The remainder of this work is structured as follows: Chapter 2 presents an overview of the macro organization of the human nervous system and outlines some of the underlying emotion-related structures.

Chapter 3 reviews the methods and techniques most commonly used in the literature regarding the assessment of psychophysiological states, while Chapter 4 brings more depth into the state-of-the-art methods for the recognition of these states using EEG signals. Current research trends on wearable sensing and virtual reality within this topic are also reviewed and some conclusions on the state-of-the-art are drawn.

Chapter 5 will focus on the development of the proposed methodology for emotion recognition in a publicly available dataset and the corresponding results.

Chapter 6 deals with the second part of the work by presenting the protocol designed for the acquisition of data in different emotion-stimulation scenarios. The results of the comparison between different visual stimuli are also present.

Finally, in Chapter 7, the main conclusions are presented, as well as the future lines of research regarding the topics that were explored throughout this dissertation.

Chapter 2

Human Nervous System and Stress Response

In order to better understand the physiological indicators of mental states, it is of great importance to first understand their cause. The first part of this chapter presents the macro organization of the human nervous system in order to provide context about some of its underlying mechanisms. Section 2.2 tries to address the question of how emotions are processed in the brain.

2.1 Human Nervous System

The human nervous system is responsible for the complex coordination of the transmission of signals between different parts of the body. Two main parts can be distinguished: central and peripheral, as shown in Figure 2.1. The central nervous system (CNS) is comprised of the brain and spinal cord, while the peripheral nervous system (PNS) includes the nerves that carry impulses to and from the CNS. (Patton, 2015)

The autonomic nervous system (ANS) is the subdivision of the PNS that regulates the physiological processes involved in functions such as breathing, heart rate, digestion, as well as the hormonal system (Furness, 2006). The ANS itself has two branches: the sympathetic and parasympathetic. These often function in an opposing manner, in order to maintain homeostasis. The parasympathetic system is mostly responsible for glands and gastrointestinal functions when the body is at rest, in a more relaxed state (Everly and Lating, 2013). On the other hand, the sympathetic system normally works to produce effective responses to specific stimuli, such as activating sweat secretion or pupil dilation. It is also responsible for the reflex regulation of the cardiovascular system. Nevertheless, under a stressful situation there is a global activation of the sympathetic nervous system, resulting in a generalized arousal, called the fight-or-flight response, whose overall effect is to prepare the individual for imminent danger (Lentz et al., 2017). This involves the activation of regulatory centres in the CNS that stimulate the hypothalamic-pituitary-adrenal (HPA) axis, one of the major systems involved in the stress response (Bao et al., 2008).

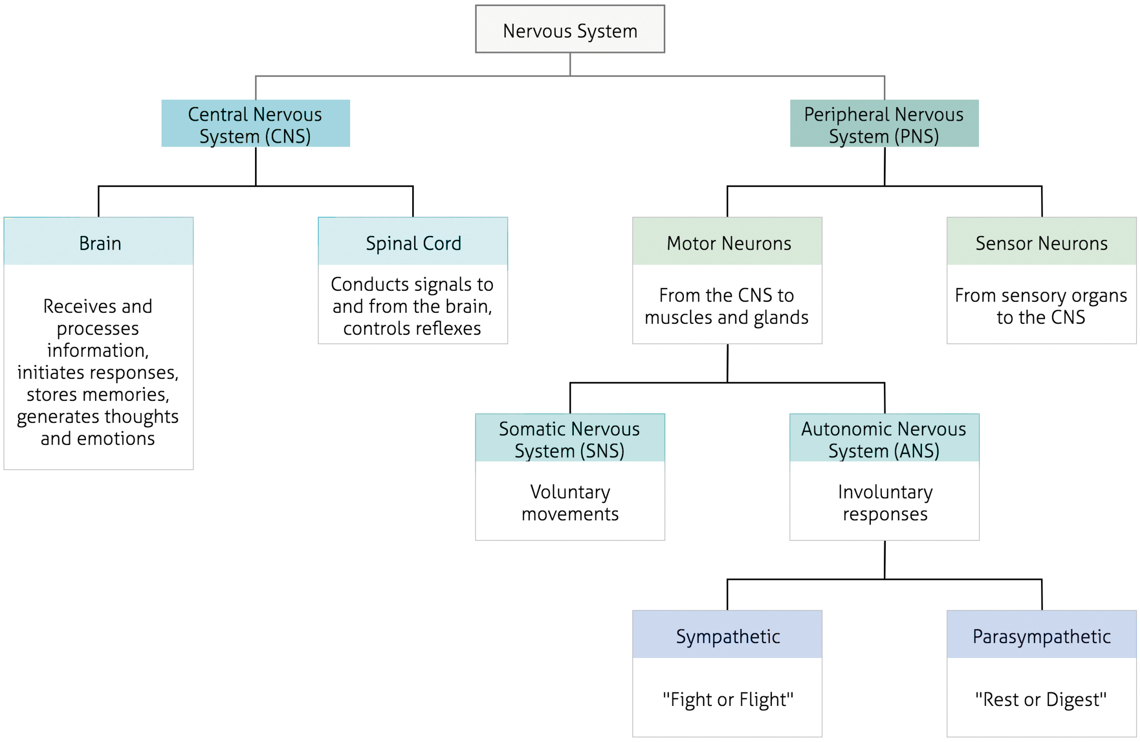


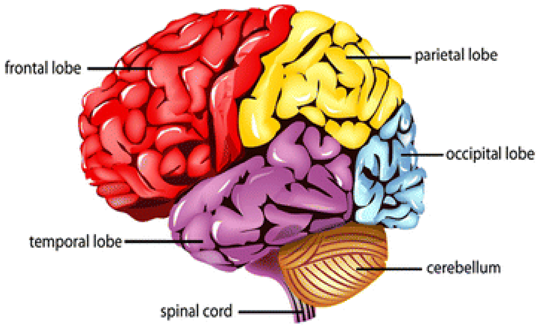
Figure 2.1: Diagram of the organization of the Nervous System.

2.1.1 Organization of the Brain

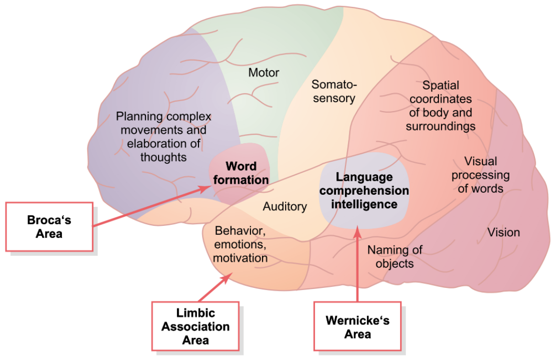
The cortex is the largest portion of the human brain and it is divided into four lobes: frontal, temporal, parietal, and occipital, as depicted in Figure 2.2a. Each of them is associated with distinct functions, distributed by specific regions, as in Figure 2.2b (Ackerman et al., 1992). In general, the frontal lobe is responsible for the conscious thoughts, including problem solving and judgement and also for motor function. The temporal lobe is involved with both memory and hearing, and the processing of complex stimuli such as faces and scenes. The parietal lobe is responsible for the integration of sensory information from various parts of the body and also takes part in the manipulation of objects. Its function also includes processing information related to the sense of touch. Finally, the occipital lobe is responsible for the sense of sight, as well as integrating visual memories with other sensations.

2.2 The Brain and Emotions

Understanding how emotions are expressed and controlled in the brain has long been a controversial and still unresolved problem among neuroscientists (LeDoux, 2000; Dalgleish, 2004). Over the decades, the field of affective neuroscience has tried to answer all kinds of questions at the interface between psychology and neurobiology and several theories arose. One of the most well-spread concepts is that of the limbic system, proposed by MacLean (1955). The limbic system is a collection of brain structures, located in the middle of the brain. Its main components are



(a) Four lobe division of the human brain. From Alotaiby et al. (2015).



(b) Map of specific functional areas in the cerebral cortex. From Hall (2015).

Figure 2.2: Functional anatomy of the brain.

represented in Figure 2.3. It has long been held responsible for controlling emotional behaviour and motivational drives and is frequently conceptualized as the "emotional brain" (Hall, 2015).

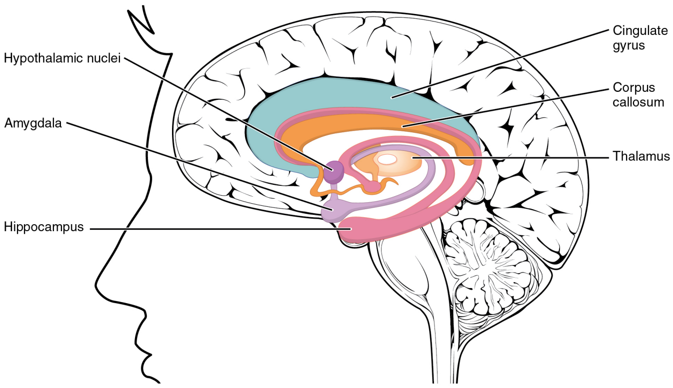


Figure 2.3: The limbic system. From OpenStax, Anatomy & Physiology. OpenStax CNX.

However, the concept of the limbic system has been subject to criticism and pointed as flawed and incomplete (Dalgleish, 2004). Many other theories have tried to relate emotions and brain structures. One of the single structures most associated with emotion processing is the amygdala. There has been proof that monkeys with damaged amygdala were “fearless” and that lesions in the human amygdala can contribute to impairments in the perception and expression of emotions and deficits in memory for emotional events Cardinal et al. (2002). Nonetheless, one of the main lines of thought now shared by researchers is that emotions may be the result of distributed networks that cross both cortical and subcortical brain regions (Kragel and LaBar, 2016; Pessoa, 2018).

Independently of how and where emotions are processed, they will trigger several important physiological events, which can be measured more or less downstream. Among them are increased oxygenation and nutrition of the brain, heart and skeletal muscles, which are all organs crucial to the “fight or flight” reaction (Chrousos, 2009). The biosignals originating from these responses, and how they can be used to assess psychophysiological states will be part of the subject of the next chapter.

Chapter 3

Psychophysiological State Assessment

The activity of the structures in the previous chapter is associated with psychological, behavioural and physiological symptoms. All of them are an expression of the subject's mental state and the analysis of such symptoms brings the possibility of inferring about it. This chapter presents a review on the methods used in the literature concerning the assessment of psychophysiological states. The first section describes the most commonly used techniques to evoke different mental states. Section 3.2 aims at the psychological evaluation of emotional states. Section 3.3 introduces some of the most common physical and behavioural measures for the assessment of those states and finally Section 3.4 reports the physiological signals used for the same purpose.

3.1 Elicitation

Carrying out mental states related research, usually means it is necessary to provoke the desired response on the subject at the required moment. Different modalities have been used for emotion elicitation including visual, auditory, tactile, or odour stimulation.

For emotion elicitation, common stimuli include sounds, music clips, film clips or images with a high emotional charge. There are several datasets available for researchers, which allow for a standardisation of experimental procedures and for the reproducibility of the results. Among them, the most widely known is International Affective Picture System (IAPS) (Lang et al., 2008). It contains 956 emotionally evocative images evaluated by several American participants on two dimensions (arousal and valence). The International Affective Digitized Sound (IADS) is a similar collection which provides a set of acoustic emotional stimuli and corresponding ratings (Bradley and Lang, 2007). Short videos or movie clips are also frequently used to evoke target emotions (Valenzi et al., 2014; Gabert-Quillen et al., 2015).

Some procedures are specifically used to induce a stressed mental state. The most commonly used techniques are the Stroop colour-word test (SCWT) (Stroop, 1935), the Trier Social Stress Test (TSST) proposed by Kirschbaum et al. (1993) and the Cold Pressor Test (CPT), first described by Hines and Brown (1936). In the most used version of SCWT, there are three tables which are supposed to be read as fast as possible. The first two represent the "congruous condition". The first requires participants to read names of colours printed in black ink and the second to name different colour patches. Conversely, in the third table the words are printed in an inconsistent colour ink (for instance the word "red" is printed in green ink) and subjects are required to name either the printed word or the colour. The SCWT has been applied by several authors in order to elicit different stress levels when developing stress recognition systems (Calibo et al., 2013; Hou et al., 2015; Jun and Smitha, 2016). The TSST is designed to exploit the vulnerability of the stress response to socially evaluative situations. The test involves 15 minutes of psychosocial stress induced by a mock job interview and followed by a mental arithmetic challenge before a panel of three judges. The discomfort associated with performance in a social context induces stress that can be measured using physiologic and/or psychological parameters. The CPT is designed to elicit pain and physiological responses associated with both HPA axis and autonomic activity (Skoluda et al., 2015). The subjects are instructed to place their non-dominant hand up to the wrist in a box filled with ice-cold water for as long as possible. Other tests such as mental arithmetic tasks (Wijsman et al., 2013; Jun and Smitha, 2016) or puzzles (Wei, 2013) are also used to induce different levels of stress. Driving simulations are also used as a stress-eliciting task (Healey and Picard, 2005).

3.2 Psychological Evaluation

It is frequent to measure perceived stress or emotional response after exposing an individual to any of the aforementioned methods. Self-report questionnaires are one of the most widely used tools to measure stress levels in humans. The Perceived Stress Scale (PSS), proposed by Cohen et al. (1983), the Stress Response Inventory (SRI) by Koh et al. (2001) and Levenstein et al.'s Perceived Stress Questionnaire (PSQ), are amongst the most common examples. The original PSS questionnaire consists of 14 questions about feelings and thoughts during the last month. The subject has to answer to these questions indicating how often he or she felt or thought that way: never, almost never, sometimes, fairly often or very often. The PSQ has a similar approach. There are 30 sentences and only 4 possible answers, describing how often the sentence applied to the subject in the last year or two. The SRI was created to score mental and physical symptoms occurred during the past two weeks. It is composed of seven stress factors that may influence the status of mental stress levels: Tension, Aggression, Somatization, Anger, Depression, Fatigue and Frustration.

In emotion-related studies the Self Assessment Manikin (SAM) (Bradley and Lang, 1994) is also widely used. It consists of a picture-oriented questionnaire, shown in Figure 3.1, containing five images for each of the three affective dimensions: valence (ranging from unpleasant/stressed to happy/elated); perceived arousal (from uninterested/bored to excited/alert); and perceptions of dominance/control. The participants rates their experience on a scale, normally with 9 points. In most cases, only the valence and arousal scales are used, since the concept of dominance is harder to be understood and expressed.

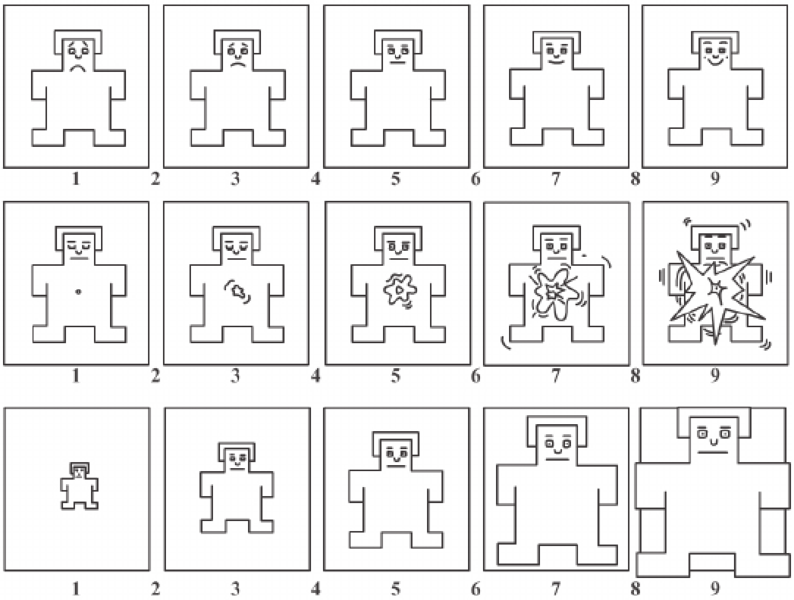


Figure 3.1: Self-Assessment Manikin (SAM) proposed by Bradley and Lang (1994). Self-evaluation scales for the dimensions of valence (top row), arousal (middle row), and dominance (bottom row) in a 9-point scale.

Another tool is the State-Trait Anxiety Inventory (STAI) designed by Spielberger et al. (1970). Form Y, the most popular version according to the American Psychological Association (2015), has 20 items for assessing trait anxiety and 20 for state anxiety. State anxiety items include declarations such as “I am worried” or “I feel calm”. In contrast, trait anxiety items are, for example: “I worry too much over something that really doesn’t matter”. All items are rated on a 4-point scale.

These questionnaire-based measurements can lead to incorrect and subjective information. By only offering information about the current emotional state, leaving unattended information about the evolution of the response with time, these tests are not suitable for monitoring and studying subtle changes that could be an indicator of a major problem (Alberdi et al., 2016). Also, these methods are intrinsically subjective and require major subject intervention. However, the lack of knowledge and insufficient capability of existing techniques impels current research to still rely on these subjective conventional methods. In most emotion recognition studies, self-assessment tests and questionnaires are used as ground truth data to evaluate the performance of the proposed systems (Alberdi et al., 2016). The widespread use of these subjective measurements alerts us that more work needs to be done to bring objective methods for measuring stress up to par (Sharma and Gedeon, 2012).

3.3 Behavioural and Physical Measures

Emotions are known to affect an individuals’ behaviour. Despite being well-known, some changes are difficult to be measured, as is the case of increased irritability and anger. However, measuring some other behavioural and physical responses is possible in an unobtrusive way and, for some cases, without the need of special equipment (Sharma and Gedeon, 2012; Alberdi et al., 2016). Examples depend significantly on the application and subject of study. Some of these features are discussed in the next sections.

3.3.1 Keystroke and mouse dynamics

In office and work-related stress monitoring, it is common to measure keystroke and mouse dynamics (Epp et al., 2011; Kolakowska, 2013). These are understood as the unique characteristics present in an individual's typing rhythm when using a keyboard or keypad or when moving or clicking a mouse, and can be an indicator of the user's stress level.

3.3.2 Posture

Another possible feature to analyse in office environments is posture. McDuff et al. (2012) have used a Kinect camera to record posture features from participants while they were at their desk, since it was proven as a good indicator about the feelings of the worker towards the tasks they are carrying out. The inclination of the person at the desk was used as an indicator of the workers' motivation. Arnrich et al. (2010) used pressure sensors installed in office chairs to analyse the changes in the posture of workers. They verified an increase in the amount of fast movements during stress tests when compared to control tests.

3.3.3 Facial Expression

Facial expression analysis is also widely used in emotion and stress recognition (Rathod et al., 2016). Dinges et al. (2005) developed an Optical Computer Recognition (OCR) technique to detect facial expressions related to workload-induced stress. They were able to discriminate between low and high stress levels using changes and asymmetry in the positioning of the eyebrows, mouth and lips. The work of McDuff et al. (2012) also used information about facial expressions as indicators of emotional valence.

3.3.4 Voice

Human voice is known to be affected by the individual's emotional state, being pitch (the fundamental frequency) and speed the most commonly observed changes (Hagmueller et al., 2006; Lu et al., 2012). Interest in speech analysis arises due to the non-intrusive nature of such measurement, providing a valuable opportunity for ubiquitous and continuous measurement of stress levels. However, quiet (e.g. a library) and noisy environments might pose some barriers to voice analysis, due to the lack of possibly acquired data and excessive noise, respectively (Adams et al., 2014; Alberdi et al., 2016). These problems mean most studies concerning stress recognition in speech usually take place in controlled environments such as the work by Fernandez and Picard (2003), who performed voice-based stress classification on a driving simulation, in a quiet environment and with a high quality microphone. Nevertheless, there are examples of stress detection both indoors and outdoors, under unconstrained settings, using smartphones for voice analysis (Lu et al., 2012; Adams et al., 2014).

3.4 Physiological Measures

As seen on Section 3.2, questionnaire-based approaches are subjective and information is conveyed in qualitative scales. Besides, this information relates with just one point in time, increasing the difficulty in monitoring a response. When using behavioural modalities for mental states detection, one of the biggest concerns is that they are prone to social masking, the possibility of individuals consciously regulating their emotional manifestations (Agrafioti et al., 2012; Selvaraj et al., 2013).

On the contrary, physiological signals allow for a real-time, objective and more realiable monitoring, presenting a promising approach for continuous assessment of psychophysiological states (Rahman et al., 2014). Being a reflection of the ANS activity, several physiological signals can provide information regarding the intensity and quality of an individual's emotional state. While the relationships between some physiological signals such as Heart Rate Variability and mental states have been proven and widely used (Seo and Lee, 2010; Castaldo et al., 2015), others need further research. The next sections present the most relevant and commonly used physiological signals.

3.4.1 Hormone Levels

When affected by stress stimuli, the organism secretes stress hormones - glucocorticoids and catecholamines such as cortisol and adrenaline (Everly and Lating, 2013; Seo and Lee, 2010). These can be measured in blood, urine or saliva. Cortisol levels follow a pronounced circadian cycle in healthy people, characterised by peak values in the morning, decreasing during the day and reaching the lowest values at night. Under stress, the ability to regulate this mechanism is affected. This means people suffering from chronic stress have elevated cortisol levels throughout the entire day (Lupien and Seguin, 2013).

Measuring salivary cortisol levels can be done in a relaxed environment, contrarily to taking blood samples, and without medical personnel. However, as with psychological questionnaires, despite being commonly used (Ham et al., 2017; Seo and Lee, 2010; Dinges et al., 2005), this method is not practical for carrying out a continuous monitoring of stress levels. Also, variables such as estrogens (related to gender, menstrual cycle, oral contraceptives) or medical conditions can affect cortisol binding and the HPA axis responsivity (Hellhammer et al., 2009).

3.4.2 Cardiovascular Activity

As stated in Chapter 2, the HPA axis and the ANS are generally considered to be the two key players in the stress response. The cardiovascular system can be used as a *proxy* for the ANS, since the release of hormones from the HPA axis results in an increase in heart rate and blood pressure

(Nicolaides et al., 2015). In addition, the Sympathetic branch of the ANS keeps the body in a more prepared and focused state with a monotonic regulation of the heart, by modulating the activity of the sinoatrial node (SA node). The SA node, which can be seen on Figure 3.2, is responsible for the contraction of the atria and, consequently, for the rhythm of heart contractions. Therefore, it is possible to infer about the ANS activity - and associated mental states - by examining cardiovascular parameters.

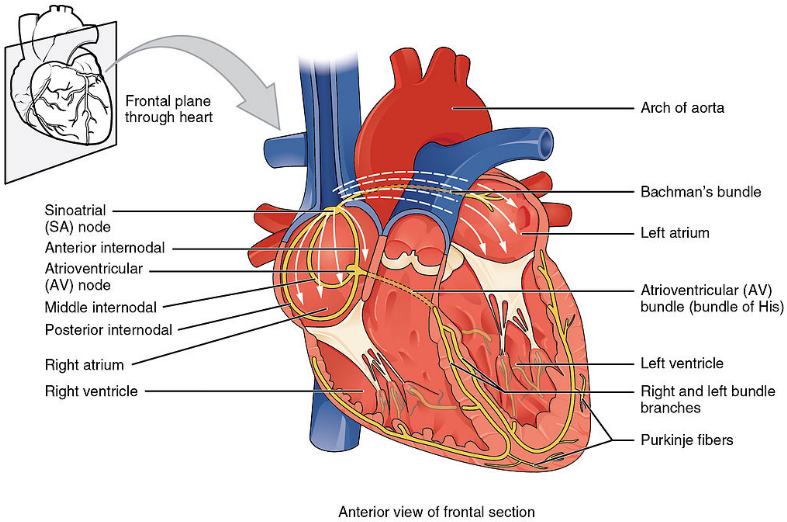


Figure 3.2: Conduction system of the Heart. From OpenStax, Anatomy & Physiology. OpenStax CNX.

3.4.2.1 Electrocardiogram (ECG) and Heart Rate Variability (HRV)

The ECG is a graphical recording of the electrical activity of the heart, responsible for the coordination of its contraction. The typical ECG waveform is represented in Figure 3.3. The main electrical signals are produced by the depolarisation and repolarisation of cardiac cells. Depolarising occurs due to the flow of ions accompanying atrial contraction, which results in a P wave. The P wave is followed shortly thereafter by the QRS complex, reflecting ventricular contraction and demarcating the onset of the systole. After the depolarisation, the ventricles repolarise, resulting in a T wave (Cacioppo et al., 2007).

HRV refers to the beat-to-beat alterations in heart rate. HRV can be analysed both in the temporal and in the frequency domain. In the first case, statistical and geometrical parameters are computed from the points and intervals represented in Figure 3.3. The frequency domain reflects the activity of the sympathetic and parasympathetic components of the ANS, in the low (LF) and high frequency (HF) power bands respectively (Bansal et al., 2009). The LF/HF ratio is the most widely used method for analysing HRV, along with the simpler time-domain features. HRV is probably the most commonly used feature in stress detection (Healey and Picard, 2005; Sharma and Gedeon, 2012; Alberdi et al., 2016; Ham et al., 2017). More recently, changes in ECG

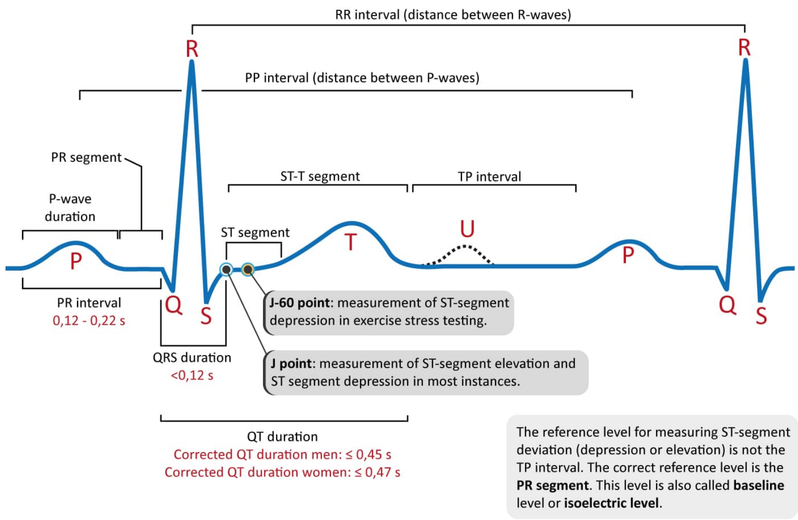


Figure 3.3: Classical ECG curve with its common waveforms, intervals and points. From Clinical ECG interpretation e-book.

morphology have also been considered as a valid indicator of mental stress (Bhide et al., 2016; Paiva et al., 2016).

3.4.2.2 Blood Pressure (BP)

BP is the pressure exerted by blood against the inner walls of the blood vessels. The use of BP for mental state assessment is, to some extent, contradictory. Kaklauskas et al. (2011) describe evidence for the dependency of blood pressure, heart rate, skin conductance and temperature on experienced stress and emotions. Nevertheless, it is reported that blood pressure is not as suitable as other physiological measurements for detecting subtle stress responses in real-time. In fact, when compared to HRV it differs substantially by the fact that, unlike HRV, blood pressure is regulated peripherally and is influenced by local conditions which can mask the changes in mental state (Hjortskov et al., 2004).

3.4.3 Brain Activity

The association of stress and negative emotions with the ANS, intuitively makes brain activity a natural source for information about mental states. However, it is not as widely used as one would expect, due to its complex nature and challenging analysis, as brain activity is characterized by significant inter- and intra-subject variability (Müller et al., 2008).

3.4.3.1 Electroencephalogram (EEG)

An EEG measures the electrical activity of the brain. It is recorded by placing an array of electrodes on the subject's scalp, according to the underlying area of cerebral cortex.

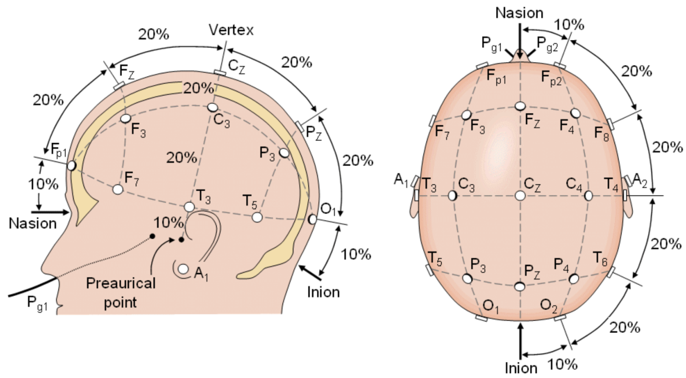


Figure 3.4: 10-20 System for electrode placement. From Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetis e-book.

EEG recording - The 10-20 system: There is a very wide spectrum of EEG devices which have been used in medical and academic research (Al-Nafjan et al., 2017). These devices range from bulky, expensive, medical-grade equipment to more portable and low-cost wearable devices, addressed in section 4.3, which have been increasingly used in EEG-related research, including for emotion recognition. Independently of which device is used, electrode placement on the scalp follows a standard to ensure cross-study reproducibility. The numbers 10 and 20 refer to the distances between adjacent electrodes, which are either 10% or 20% of the total front–to-back or right-to–left distance of the skull. In this system each electrode has a code to identify the lobe and hemisphere location. The letters F, T, P and O represent the brain lobes. C stands for a central position. Even numbers refer to the right hemisphere, while odd numbers indicate the left hemisphere. Additionally, the letter codes A, Pg and Fp identify the earlobes, nasopharyngeal and frontal polar sites, respectively (Jasper, 1958).

Despite its poor spatial resolution and the need for many electrodes, the EEG provides a great time resolution, which allows for the study of subtle changes in response to emotional stimuli (Alarcao and Fonseca, 2017).

EEG signals are usually divided into four main frequency bands: Alpha (8–13 Hz), Beta (13–30 Hz), Delta (0.1–4 Hz) and Theta (4–8 Hz). Alpha waves reflect a calm, open and balanced psychological state, decreasing in stress situations. Beta activity reflects cognitive and emotional processes, so it increases with mental workload and stress (Rowland et al., 1985; Matlovič, 2016). Theta activity is normally seen in the parietal and temporal regions in children, but they also appear during emotional stress in some adults, particularly during disappointment and frustration. Delta waves are associated with unconscious states, such as deep sleep stages Hall (2015). Seo and Lee (2010) analysed the relationship between EEG, ECG and cortisol levels when measuring chronic stress. They verified a significant positive correlation between the cortisol levels and the high Beta activity at the anterior temporal sites. Besides frequency-domain features, researchers have used statistical measures of EEG activity, as well as more complex features such as Fractal Dimension or connectivity measures (Hosseini and Khalilzadeh, 2010; Valenzi et al., 2014; Subhani et al., 2017).

3.4.3.2 Functional Magnetic Resonance Imaging (fMRI)

fMRI is an imaging technique that is capable of measuring brain activity. Brain activity in a particular region increases the local oxygen consumption. To meet this increased demand, the blood-flow to that region increases. fMRI is able to detect these changes in blood oxygenation by means of a contrasting agent. The activation maps obtained by fMRI are a way of showing which parts of the brain are involved in a particular mental process. Hayashi et al. (2012) used fMRI to investigate whether stress states affect brain responses in the regions related to emotional and cognitive processing. Their results showed no significant differences on brain regions related to emotional processing between the stressed and non-stressed groups. However, they reported less activity in cognitive processing brain regions on the stressed group.

3.4.3.3 Near-infrared Spectroscopy (NIRS)

NIRS is an additional source of information about brain activity. It is a relatively new modality which measures oxygenated and deoxygenated haemoglobins in the superficial layers of the human cortex. It works by sending near-infrared light into the head and measuring the light attenuation resulting from photon interactions in the cortical tissues (Al-Shargie et al., 2016). This attenuation is converted into concentration changes in oxygenated and deoxygenated haemoglobins based on a modified version of the Beer-Lambert law (Coyle et al., 2007). The technique has gained particular interest in BCI studies due to its portability, low-cost and sensitivity (Coyle et al., 2007). In most cases it is used conjointly with the EEG for the detection of the onset and offset of motor imagery tasks (Fazli et al., 2012; Tomita et al., 2014). Nevertheless, it has also been used by Al-Shargie et al. (2016) for mental stress assessment. They observed a decline in oxygenated haemoglobin during the stress-inducing mental arithmetic task. They also argue that the fusion of EEG and NIRS features improves the accuracy of mental stress detection.

3.4.4 Peripheral Measures

3.4.4.1 Electrodermal Activity (EDA)

The EDA, also known as Electrodermal Response, Galvanic Skin Response, Skin Conductance Activity or Skin Conductance Response, is defined as a change in the electrical properties of the skin. When an individual is subject to emotional arousal, increased cognitive workload or physical activity, the level of sweating increases, which changes the properties of the skin, increasing conductance and decreasing resistance (Critchley, 2002).

EDA has been widely used in stress and emotion detection as can be observed in the extensive literature review by Alberdi et al. (2016). It is considered one of the best real-time correlates of stress and is sometimes used as ground truth for other signals (Hernandez et al., 2014; Healey and Picard, 2005).

3.4.4.2 Electromyogram

An Electromyogram (EMG) measures the electrical activity of the muscles by using electrodes placed over the muscle of interest. Wijsman et al. (2013) verified that mental stress provoked a significant increase in the trapezius muscle activity, translated into an increase in EMG amplitude and a decrease in periods of relaxation. Healey and Picard (2005) also recorded EMG from the trapezius muscle in subjects doing a stress-inducing real-life driving task. Facial EMG has also been used to detect affective states (Van Boxtel, 2010). However, both techniques can be obtrusive and difficult to apply in practical and realistic scenarios.

3.4.4.3 Pupil Diameter, Blinking and Eye Gaze

The development of infrared eye tracking systems, along with Image Processing techniques, has made possible the measurement of pupil diameter, eye gaze and blink rates. Since pupil dilation is regulated by the ANS (Cacioppo et al., 2007), it has been used to infer emotional states (Haapalainen et al., 2010). However, both negative and positive stimuli can cause pupil diameters to change, increasing the difficulty of using this measurement (Sharma and Gedeon, 2012). Eye gaze enables deducing the individual's mental states and intentions by providing information on his attention source. Blinking rate is also sensitive to stress. Nevertheless, there are contradictory results reported in the studies using these measures (Sharma and Gedeon, 2012).

Chapter 4

EEG-based Mental State Recognition: Datasets, Methods and Technologies

As previously seen on section 3.4, there are numerous physiological signals that are a convenient and reliable means for the evaluation of an individual's mental state. Among them, the use of EEG signal emerges as an obvious source, given the apparently more direct link to the underlying cognitive processes. However, the complexity of the signal and its acquisition has hindered EEG-based emotion detection as a noticeable and legitimate research topic. Nevertheless, recent times have witnessed a growth in EEG-based emotion detection publications, as pointed by Al-Nafjan et al. (2017). This growth has been spurred by the scientific and technological advancements contributing to the proliferation of EEG devices and to the ever-improving machine learning techniques. Thus, this chapter aims at addressing mental state recognition using the EEG signal. Section 4.1 attends to some of the datasets used as benchmarks in emotion recognition-related research. Section 4.2 presents a brief explanation of the main methods currently used in this field, as well as some of the most relevant works. Sections 4.3 and 4.4 address two of the most recent trends on the topic, namely the use of wearable, low-cost devices and virtual reality environments for psychophysiological assessment. Lastly, in Section 4.5 some final remarks are drawn.

4.1 Datasets

The effects on the subjects' emotional state are obviously influenced by the quality of the presented stimuli. Therefore, analysing a benchmark dataset is of vital importance in order to obtain meaningful and comparable results. Recent advances in emotion recognition have motivated the creation of novel databases containing emotional expressions in different modalities and in response to well-defined stimuli.

The DEAP dataset, developed by Koelstra et al. (2012), is the benchmark dataset for emotion analysis (Alarcao and Fonseca, 2017). It comprises EEG and 13 peripheral physiological signals for 32 subjects who individually watched 40 one-minute music videos of different genres as a stimulus to induce different affective states. For EEG recording a Biosemi ActiveTwo system was used at a sampling rate of 512 Hz. Participants rated each video in terms of arousal, valence, like/dislike, dominance and familiarity, using the SAM 9-point scale. Frontal face video was also recorded in 22 participants.

Being very similar to the DEAP dataset, the MAHNOB-HCI dataset (Soleymani et al., 2012) is also commonly used in emotion recognition. In this case 30 participants were shown 20 fragments of movies and pictures. Physiological signals including ECG, EEG (32 channels), respiration amplitude, and skin temperature were recorded while the videos were shown to the participants. Facial expression, audio signals and eye gaze were also collected.

The SEED dataset (Zheng et al., 2017) is another dataset for emotion recognition using EEG signals. Facial videos and EEG data were recorded from 15 subjects while watching 15 Chinese film clips (positive, neutral and negative emotions). The EEG was recorded using an ESI Neuro-Scan System with 62 channels, at a sampling rate of 1000 Hz.

The aforementioned databases required specialised and costly equipment in order to capture the biosignals. As a result, the application of the proposed algorithms becomes confined to controlled environments, suitable for such devices, excluding everyday scenarios that could benefit from affective computing. In an attempt to address this limitation, Katsigiannis and Ramzan (2018) have very recently created a database for emotion recognition through EEG and ECG Signals using wireless, low-cost and off-the-shelf devices. The DREAMER dataset includes EEG and ECG recordings from 23 subjects who watched 18 film clips selected and evaluated by Gabert-Quillen et al. (2015). The EEG was recorded at a sampling rate of 128 Hz using the Emotiv EPOC system and the ECG data was obtained with the SHIMMER™ wireless sensor. Subject's evaluated their emotional states by reporting valence, arousal and dominance in a 5-point SAM.

4.2 Methods for Emotion Recognition with EEG signal

Notwithstanding its complexity, several authors have investigated the use of EEG for recognizing user mental states. The majority of these studies aim to detect basic sets of emotions (Valenzi et al., 2014; Matlovič, 2016; Mehmood and Lee, 2016) or simply distinguish between positive, negative, and neutral emotions, based on the valence-arousal dimensional model (Hosseini and Khalilzadeh, 2010; Menezes et al., 2017). Other works focus solely on stress recognition (Hou et al., 2015; Jun and Smitha, 2016; Subhani et al., 2017).

The development of an emotion recognition system from physiological signals in general, and from EEG in particular, follows a typical machine-learning pipeline. After the acquisition of the signal (while the user is being exposed to the stimulus being tested), the recorded signals are pre-processed so as to reduce noise and artifacts. Then, the relevant features are extracted and selected. A classifier is trained based on a training set and using those computed features, leading to the interpretation of the original brain signals. The following sections try address the most commonly used methods in EEG-based emotion recognition in a more sistematic way. Table 4.1 summarizes some of the most relevant and recent works discussed regarding emotion recognition from EEG signal.

Table 4.1: Summary of most relevant literature reviewed.

Reference	Target emotion(s)	Elicitation	No. of subjects	Recording device	Features	Decision	Accuracy (scheme)
Hosseini and Khalilzadeh (2010)	Calmness, negative excitement	IAPS	15	FlexComp Infiniti	Fractal Dimension (Higuchi) Correlation Dimension Power (DWT): α, β, γ, δ and θ	Elman Network	82,7% (2-class)
Valenzi et al. (2014)	Sadness, disgust, neutral, amusement	Videos	9	Biosemi Active 2	Power (STFT): α, β , γ, δ and θ	SVM (best)	97,2% (4-class)
Liu and Sourina (2016)	Satisfied, happy, surprised, protected, sad, unconcerned, angry, frightened	IADS IAPS	14 (IAPS) 16 (IADS)	Emotiv EPOC+	Statistical Fractal Dimension (Higuchi) HOC	SVM	IADS: 87,02% IAPS: 76.53% DEAP: 90,35% (2-class)
Mehmood (2016)	Happy, calm, sad, scared	IAPS	44	Emotiv EPOC+	Statistical Frequency Power (LPP-based)	SVM k-NN (k=5)	58% (4-class)
Matlovic (2016)	joy, sadness, anger, surprise, disgust, fear, neutral	Videos (DEAP)	9	Emotiv EPOC+	Power (DWT): α and β Valence/arousal representations	SVM	36% (7-class)
Shin et al. (2017)	amusement, fear, sadness, joy, anger, disgust	Videos	30	Laxtha QEEG-8	Power (FFT): α, β and θ HRV features (LF/HF)	MLP SVM NB	98,1% (6-class) (EEG only: 62.3%)
Menezes et al. (2017)	Valence / arousal ratings	-	-	-	Statistical PSD HOC	SVM RF	DEAP: Valence: 88,4% Arousal: 74% (2-class)
Nakisa et al. (2018)	LA-P, LA-N, HA-P, HA-N	Videos	13	Emotiv Insight	Statistical Fractal Dimension (Higuchi) Power (DWT, PSD): α, β, γ, δ, θ HOC Hjorth parameters	PNN	65% DEAP: 67,47% MAHNOBI: 97,11% (4-class)
Li et al. (2018)	Positive, negative	-	-	-	Time-frequency Non-linear dynamical features	SVM	DEAP: 59,06% SEED: 83,33%

4.2.1 Stimuli

Different modalities can be used for emotion stimulation, as presented on Section 3.1. These emotional stimuli are usually selected to cover the desired valence and arousal levels. Nevertheless, the majority of studies which measure emotion from EEG signals use pictures from the IAPS (Alarcao and Fonseca, 2017). This is the case of Hosseini and Khalilzadeh (2010), who used a subset of this database to elicit two target emotions: calmness and negative excitement. In their protocol 15 participants sat in front of a computer screen which displayed 8 blocks of 4 images (in order to ensure stability of the emotion over some time). Before each new block, a dark screen with an asterisk in the middle was shown for 10 seconds. This intended to make a clear separation on the subjects' emotional state and draw their attention. Liu and Sourina (2016) used both audio (IADS) and visual (IAPS) stimuli to induce a range of emotions in 14 and 16 participants, respectively. Five sound clips and four images were used for each emotion. More recently, Mehmood and Lee (2016) also used pictures form the IAPS database to evoke emotions. They selected 45 pictures for each emotional state (sad, scared, happy, and calm). These pictures were randomly shown to young students for 1.5 seconds, interspersed with 0.5 s intervals showing a black image.

Similarly to presenting emotional images, sometimes videos or film clips are used. This was the approach followed by Valenzi et al. (2014) to study sadness, disgust, amusement and neutral emotions. Matlovič (2016) also used emotional videos so as to study 6 basic emotions. In this case, the same music videos from the DEAP database were used. This allowed for a direct comparison of the results from their experiment to the one proposed by the authors of the DEAP study (Koelstra et al., 2012).

4.2.2 Signal pre-processing

The EEG signal is characterized by being intrinsically noisy, due to the very weak electrical activity produced by the brain, which is in the order of microvolts. Therefore, pre-processing the EEG signal for noise and artifact removal is crucial. The most common approach is to use a bandpass filter, since frequencies above 30 Hz are usually not interesting for EEG analysis. The cut-off frequencies vary slightly across the different works. Hosseini and Khalilzadeh (2010), used a 0.5-35Hz bandpass filter. Valenzi et al. (2014) filtered the EEG signals in the range of 0.16-70Hz and used a 50 Hz notch filter to remove power line noise. Nakisa et al. (2018), also used a notch filter, but opted for using a 4-64Hz bandpass filter in the first place. However, according to the survey by Alarcao and Fonseca (2017) the most common frequency range is 4-45 Hz. In most studies eye-blinking and movement artefacts are manually excluded by eliminating the corresponding part of the recording. Common automatic techniques include Blind Source Separation and Independent Component Analysis (Alarcao and Fonseca, 2017). It is also common to re-reference the electrodes with a Common Average Reference.

4.2.3 EEG Features

In emotion recognition from EEG there is no consensus upon which features are the most appropriate (Jenke et al., 2014; Menezes et al., 2017). This results in a multitude of possible data to extract from the EEG signal, as verified in Table 4.1. In general, features are separately extracted from each of the recording channels. For a better understanding of the features used in the previously mentioned works, this section will try to briefly address some of the most frequently used ones, by dividing them in time-domain, frequency-domain and connectivity features.

4.2.3.1 Time-domain Features

Many authors explore six different statistical features for all the bands of each channel for their simplicity and reasonable results (Jenke et al., 2014; Menezes et al., 2017; Li et al., 2018):

(a) Mean (μ_x);

(b) Standard deviation (σ_x);

(c) Mean of the absolute values of the first differences (δ_x):

$$\delta_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |x[n+1] - x[n]| \tag{4.1}$$

(d) Mean of the normalised absolute values of the first differences ($\overline{\delta_x}$):

$$\overline{\delta_x} = \frac{\delta_x}{\sigma_x} \tag{4.2}$$

(e) Mean of the absolute values of the second differences (γ_x):

$$\gamma_x = \frac{1}{N-2} \sum_{n=1}^{N-2} |x[n+2] - x[n]| \tag{4.3}$$

(f) Mean of the normalised absolute values of the second differences ($\overline{\gamma_x}$):

$$\overline{\gamma_x} = \frac{\delta_x}{\sigma_x} \tag{4.4}$$

Another frequently used measure is the Fractal Dimension (FD). Fractals are objects or sequences which have a similar appearance when observed at different scales. These objects have details at arbitrarily small scales, making them too complex to be represented by Euclidean space, and so they are assigned a dimension which is non-integer (Kumar et al., 2017). This feature provides information of the space filling and self-similarity of a time series such as the EEG and is considered to be positively correlated with how energized the user is, hence relating to the arousal level. Hosseini and Khalilzadeh (2010) and Nakisa et al. (2018) extracted this feature by applying Higuchi's algorithm (Higuchi, 1988).

Menezes et al. (2017) used yet another time-based feature, known as High Order Crossings (HOC). A finite zero-mean time-series Z_t with $t = 1, ..., t = N$, oscillating about level zero will have a number of zero crossings (NZC). In general, when a filter is applied to a time series its oscillatory behaviour will change, changing its NZC, too. By applying a specific sequence of filters, a corresponding sequence of NZC is obtained. This is the HOC sequence (Petrantonakis and Hadjileontiadis, 2010).

4.2.3.2 Frequency-domain Features

Power features from the different EEG wave bands are the most notorious in the context of emotion recognition from EEG signal. The Fast Fourier Transform (FFT) is the algorithm of choice for computing the Discrete Fourier Transform (DFT) of a sequence, dividing it into its frequency components (Hou et al., 2015; Jun and Smitha, 2016; Shin et al., 2017; Subhani et al., 2017). The DFT is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \tag{4.5}$$

The FFT algorithm makes this computation faster, so that working in frequency domain is equally computationally feasible as working in temporal domain. Most works use mainly the Delta, Theta, Alpha, Beta, and Gamma bands when extracting power features from the signals.

However, there are several combinations in the literature and some authors even analyse just one band. Menezes et al. (2017) and Nakisa et al. (2018) have used an alternative method by estimating the Power Spectral Density (PSD) of each frequency band. The method is based on the concept of using periodogram spectrum estimates, which are the result of converting a signal from the time domain to the frequency domain.

Another technique is the Discrete Wavelet Transform, used by Hosseini and Khalilzadeh (2010) and Matlovič (2016), which decomposes the signal in different approximation levels corresponding to different frequency ranges. The extracted wavelet coefficients can provide a compact representation of the energy distribution of the EEG wave bands, while conserving the time information of the signal (Jenke et al., 2014). Both works have computed the average power of these coefficients. Hosseini and Khalilzadeh (2010) further implemented the mean and standard deviation of these values as features.

4.2.3.3 Connectivity Features

Differently from most other authors, Lee and Hsieh (2014) employed a set of brain connectivity measures. In their work three connectivity measures were used to estimate functional connectivity: correlation, coherence, and phase synchronization of pairs of electrodes. An increased correlation corresponds to a stronger relationship between two brain locations. Coherence relates to the degree of association between two brain locations at a specific frequency.

4.2.4 Feature Selection

The poor spatial resolution of the EEG signal forces the use of multi-channel devices so as to obtain meaningful information. EEGs are often acquired with 32, 64 or even more electrodes. This presents a problem to classification and recognition systems. In fact, additional EEG channels usually include noisy and redundant features, which not only increase computational complexity, but also degrade performance (Zhang et al., 2016). To tackle this problem, feature selection methods can be applied. Examples include ReliefF, Min-Redundancy-Max-Relevance (mRMR) (Jenke et al., 2014), evolutionary computation algorithms (Nakisa et al., 2018), ANOVA (Lee and Hsieh, 2014) and Linear Discriminant Analysis (LDA) (Valenzi et al., 2014). However,some works in the literature neglect this important step or use simplistic approaches such as manually eliminating some channels (Menezes et al., 2017; Matlovič, 2016).

4.2.5 Classification Approach and Results

Several machine learning (ML) algorithms have been for used emotion and stress recognition from EEG signals. By far, the most frequent examples are support vector machines (SVM) and k-nearest neighbours (k-NN) (Al-Nafjan et al., 2017; Alarcao and Fonseca, 2017).

A k-NN classifier is one the simplest ML algorithms. It is a non-parametric method in which an object is classified by a majority vote of its neighbours. The object is assigned to the most common class among its k nearest neighbours. If k = 1, then the object is simply assigned to the class of that single nearest neighbour. A simple 2-dimensional example is shown in Figure 4.1.

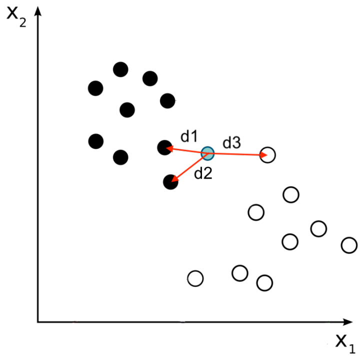


Figure 4.1: k-NN 2-dimensional example. In this case $k = 3$. The red arrows represent the distance to the 3 nearest neighbours. Assuming equal voting weights, the blue point would be assigned to the black class.

The distance to neighbouring points can be computed by different functions, being the Euclidean distance the most common one.

A SVM (Cortes and Vapnik, 1995) is a supervised learning model that constructs a hyperplane in a high-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation between classes is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class. This is called the functional margin. In general, the larger the margin the lower the generalization error of the classifier. This principle is illustrated in the 2-dimensional example in Figure 4.2.

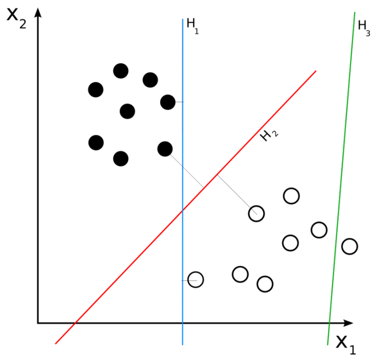


Figure 4.2: SVM 2-dimensional example showing how a a separating hyperplane would be obtained for two classes of points. H3 does not separate the classes. H1 does, but only with a small margin. H2 separates them with the maximum margin.

One of the strengths of SVMs is that they focus on the interface between two groups and not on samples that are far away from the dividing hyperplane, making them able to capture the subtle differences that contribute to distinguishing the classes.

Regardless of the use of some specific classifier, there are several approaches to build an emotion-recognition system. The model can be subject-specific or it can be intended for a subject-independent application. Furthermore, it can be designed for offline or online use (Alarcao and

Fonseca, 2017). The next paragraphs try to clarify the approaches followed by some of the authors, which can in turn shed light in the results obtained.

In order to differentiate calm and negatively excited situations, Hosseini and Khalilzadeh (2010) fed an Elman network with complex features, already discussed. An Elman network is a three-layer back-propagation neural network with an additional feedback connection from the output of the hidden layer to the input (Elman, 1990). The results of this study indicate a maximum accuracy of 82,7%.

Valenzi et al. (2014) tested several supervised learning and clustering algorithms for a subject-dependent recognition of 4 emotional classes. Their best results were obtained with a linear SVM, which yielded an average of 97,2% among 9 participants. This high performance is attributed by the authors to the high quality of the collected EEG data and controlled protocol. Liu and Sourina (2016) also used SVM, this time with a polynomial kernel, to recognise 8 emotions. Their algorithm also represents a subject-dependent approach and the best reported result is 90,35% when the algorithm was tested in the DEAP dataset. However, this result corresponds to the mean recognition for combinations of just two emotions.

Mehmood and Lee (2016) evaluated a SVM and a k-NN ($k = 5$) for discriminating 4 basic emotions. They propose a feature extraction method based on late positive potential (LPP), which is a pattern that indicates the attention paid by a subject to a visual stimuli. Their results indicate an average recognition rate at the Theta and Alpha bands of 56,2% and 57,9% for the k-NN and SVM, respectively.

The work of Matlovič (2016) aimed at recognizing 6 basic emotions. Their method, which was first validated with the DEAP dataset, extracted arousal and valence numeric representations from the Alpha and Beta power of the EEG. These representations where then used as features for a SVM. An accuracy of 36% when predicting one of seven emotions (a neutral state was added to the set) was obtained. In spite of being low, the author notes that this value is above the probability of random selection.

The work of Shin et al. (2017) used highly complex features combining the EEG and ECG signals for discriminating six basic emotions. Those features were fed to a SVM, a Multi-layer Perceptron (MLP), which is a feedforward artificial neural network, and a Bayesian Network (BN), which is a probabilistic graphical model, representing a set of variables and their conditional dependencies. The BN achieved the highest accuracy, averaged across all channels, of 98,06%. However, this value was obtained when using the complex combined features of ECG and EEG signals. For a fair comparison, it must be noted that using EEG-only features, the method yielded a maximum accuracy of 62,28%.

Menezes et al. (2017) used the EEG signals in the DEAP dataset for classifying valence and arousal states with a SVM and a Random Forest (RF) classifier. A RF works by constructing a multitude of decision trees when training. The output class is the one that is the mode of the classes. A separate analysis was done to evaluate valence and arousal. The best reported accuracies were 88,4% for valence in a two-level scheme (high vs low), achieved by both classifiers, and 74,0% for arousal, obtained by the RF. The recent work of Li et al. (2018) also used the data from the DEAP dataset, but, contrarily to other works, they focused on a cross-subject methodology. After extracting a substantially wide range of features and automatic feature selection, they used a SVM to evaluate the ability to discriminate positive and negative emotions. They report 59,06% and 83,33% as the best performance obtained in the DEAP and the SEED datasets, respectively.

More recently, Nakisa et al. (2018) chose a probabilistic neural network (PNN) to evaluate the performance of different feature selection algorithms in an emotion recognition framework. They classified four emotions, according to the valence-arousal plot (Low Arousal-Positive, Low Arousal-Negative, High Arousal-Positive and High-Arousal-Negative). In the data recorded in their study an accuracy of 65% was achieved. However, the algorithm reached the best performance in the MANHOBI dataset with 97,11% of accuracy. In the DEAP dataset the best result was 67,47%.

4.3 Low-cost EEG recording

Recent developments in technology, in particular the integration of wearable and embedded sensors, have made physiological signals acquisition a more affordable and straightforward process. Many emotion- and stress-related studies now rely on the use of convenient devices that enable physiological monitoring in more unconstrained environments. In what EEG is concerned, there is a myriad of recording devices. The more recent, wearable ones offer attractive features such as wireless transmission, low power and memory consumption, real-time filtering and de-noising, and a user-friendly experience with regards to usability and comfort (Athavale and Krishnan, 2017).

The predominance of low-cost EEG-recording devices in recent literature proves that wearable and low-cost monitoring is simultaneously a promising research tool and real-life alternative.

4.4 Virtual Reality and Psychophysiology

Virtual Reality (VR) can be understood as a collection of technologies such as 3D displays, motion tracking hardware, input devices and software frameworks, which aim to generate realistic images, sounds and other sensations that simulate the user's senses so as to simulate the presence in a virtual or imaginary environment. The advent of consumer grade hardware (such as the Oculus Rift, the Samsung Gear VR or the Google Cardboard, shown in Figure 4.3), as well as the maturity of software platforms to create and display VR contents indicates that this could be the next big wave of excitement in computer technology (Parisi, 2015). Interestingly, the Hype Cycle in Figure 1.2 (see Chapter 1), also shows that VR is already in the Slope of Enlightenment phase, with mainstream adoption predicted to occur in just 2 to 5 years.

This growing trend has also reached psychophysiology and neuroscience, usually in association with HCI. In fact, Rey and Alcañiz (2010) emphasizes the two-way relationship that has been created between virtual reality and neuroscience and the benefits that can emerge from the cooperation between these two fields. One of the examples is the use of neuroscience tools and methods, such as the analysis of EEG signals, to understand and improve the sense of presence

(a) Oculus Rift (b) Samsung Gear VR (c) Google Cardboard

Figure 4.3: Consumer grade Virtual Reality hardware.

in virtual reality environments. The work of Baumgartner et al. (2006) was the first to analyse the EEG activity during different virtual reality experiments. They examined the spatial presence of participants who sat in front of a regular computer screen and watched different virtual roller coaster scenarios, without any type of interaction. They found increased activation in the Alpha band in parietal brain areas. Later, Kober et al. (2012) confirmed that these results hold true in an environment with user-interaction. They examined the spatial presence using two different VR systems: a more immersive Single-Wall-VR-System and a less immersive Desktop-VR-system and confirmed that the single-wall system resulted in an increased activity in parietal areas. A similar study was conducted by Slobounov et al. (2015), who reported that a higher FM-Theta (Frontal midline Theta rhythm) power was associated with a higher sense of presence.

Although the previous works present valuable insights into the immersion experience in different VR environments, few authors have tried to sistematically investigate the effects of different modalities of stimuli and how it can influence the emotional response of the user. In fact, according to Uhrig et al. (2016), when comparing pictures and normal videos for emotion stimulation, although there are many theoretical arguments in favour of films, not much empirical evidence can be found. To address this issue, these authors conducted a study to compare pictures and films in their capacity to provoke emotions. In this work, 139 participants rated their experience using a paper-and-pencil version of the SAM scale. Their results indicated that the pictures with positive content resulted in lower arousal ratings, when compared to the film clips, although this effect was not consistently significant. On the other hand, they found that film clips were less effective than pictures in producing negative emotions. Dhaka and Kashyap (2017) conducted a similar investigation that yielded similar results. Despite the interesting results, both studies are limited by the statistical analysis of explicit measures in the form of valence and arousal ratings, which may not correspond to the underlying neural processes.

To the best of my knowledge, the work of Kweon et al. (2017) is the first to study the brainwave patterns of participants who were exposed to 2D (flat) and VR (stereoscopic) videos while having their EEG recorded. The content was displayed in a 40-inch TV screen (2D) and in the Oculus headset (VR). Despite the fact that the analysis of the signals is quite limited, the results indicate that Alpha wave activity is predominant in 2D videos, while Beta activity is higher in the VR videos, confirming the empirical idea that there are differences in brain processes while watching VR videos, when compared to 2D.

4.5 Remarks and Conclusions

The literature review devised in this chapter, allowed for a better understanding of the major topics in psychophysiological state assessment. The rationale behind using physiological signals for this purpose is advantageous in the sense that these can hardly be deceived by voluntary control and require no or minimal user intervention. By providing relevant parameters which can be extracted, physiological signals, and the EEG in particular, can contribute to the understanding of a full array of emotions which can be used in a multitude of applications. The use of emotions as a measure of an individual's adaptation to environmental process, as opposed to a more conservative stress-only recognition, can potentially yield better insights. Nevertheless, when analysing the state-of-the-art in emotion and stress recognition from the EEG signal in more depth, there are some indicators that there is room for development and improvements. Although most works report their results only in terms of accuracy, the performance of EEG-based emotion recognition systems are still deviated from what would be desirable in a real application. Furthermore, most works present a subject-based classification, which contradicts the ideal of developing a universal system envisioned to recognise emotions from any user. Also, some of the best results were achieved using methods that are not the most suitable if one is to expand the studies to real-time and embedded systems, where computational cost must be taken into account. Despite these limitations, great advances have been made and these can be partially explained by the surge of consumer grade hardware which has enabled a pronounced growth, both in EEG and Virtual Reality studies and applications. Virtual Reality also comes as a new breath of air in HCI, since it contributes to the engagement and greater sense of presence of the user. This fact should have an interesting relationship with the concept of emotional arousal, which is, in some cases difficult to understand and quantify.

The field of emotion recognition is a multidisciplinary research area, which requires knowledge and information from a vast landscape of scientific domains. This fact not only contributes to the myriad of different perspectives explored in the literature, but also to the significance of this topic. Given the different aspects inherent to goals of this project, the state-of-the-art methods and results will serve as a basis for the development of an EEG-based emotion recognition pipeline, as well as to gain insights into the interpretation of the signals for different stimuli.

Chapter 5

Emotional State Recognition in the DEAP Dataset

Having identified some room for improvement in the literature review, we developed a methodology to identify different emotional states in the publicly available DEAP dataset. This is the most used dataset in the literature (Alarcao and Fonseca, 2017), allowing for a more accurate validation of the results. The main goal was to follow a traditional machine learning pipeline based on the state-of-art methods and improve some of the steps, hopefully leading to an improvement in the final performance. A schematic overview of the proposed pipeline is illustrated in Figure 5.1.

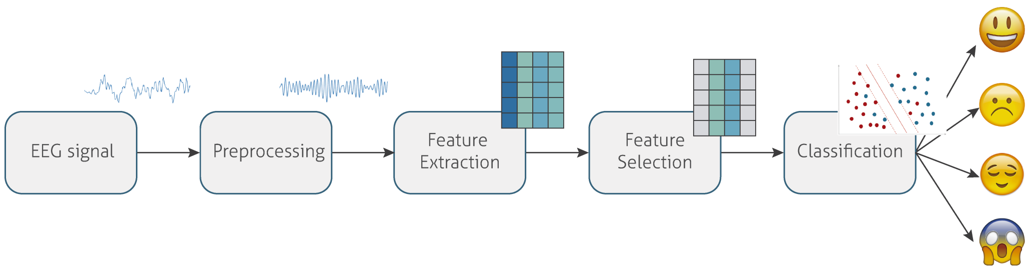


Figure 5.1: Typical machine learning pipeline for EEG emotion recognition.

The next section details the preprocessing of the available EEG signals. Sections 5.2 and 5.3 address the extracted features and the corresponding feature selection. Section 5.4 explains the machine learning algorithms used for classification. Section 5.5 discusses the results obtained and finally in Section 5.6 some general remarks about the results are drawn.

5.1 Data preprocessing

As described in Section 4.1, the DEAP dataset contains EEG signals with 32 channels, recorded from 32 participants while watching 40 music videos, each with one minute. The participants rated the videos based on the Arousal-Valence, like/dislike, and dominance. In order to decrease computational time, only 24 videos were considered in this work. These were grouped in four

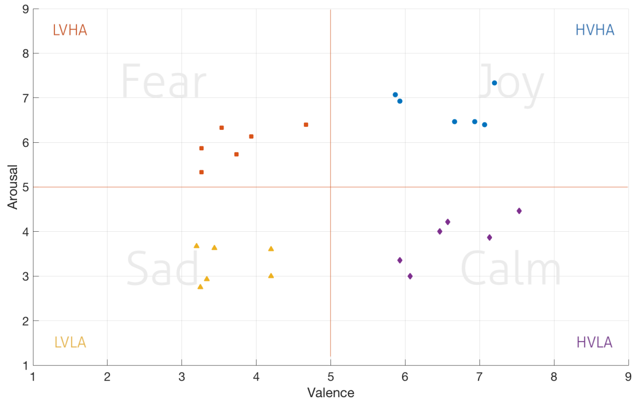


Figure 5.2: Ratings of the chosen videos from DEAP.

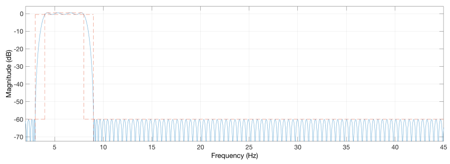
emotional classes (joy, calmness, sadness, fear) according to their average arousal and valence ratings, as shown in Figure 5.2. For each subject, the EEG signals corresponding to these 24 videos were analysed so as to extract relevant features that could feed a classifier.

The DEAP EEG signals are already preprocessed so as to remove noise and artifacts. According to the dataset description webpage (DEAP, 2012), the original data was downsampled to 128 Hz and the artifacts resulting from eye movements were removed with a blind source separation technique, as described in Koelstra et al. (2012). In order to keep the frequencies of interest, a bandpass frequency filter from 4.0-45.0 Hz was applied, and finally the data was averaged to the common reference.

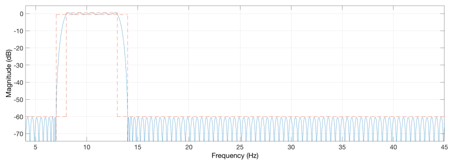
5.1.1 Bandwave filtering

For each of the 32 channels, the Alpha, Beta and Theta rhythms were extracted by using appropriate bandpass filters, according to the frequency ranges of each brainwave. Since the DEAP data had already been filtered, the Delta waves were not considered in this work. Equiripple Finite Impulse Response (FIR) filters were designed using the Parks–McClellan algorithm (McClellan et al., 1973), according to the MATLAB implementation. The magnitude response of these filters is represented in Figure 5.3.

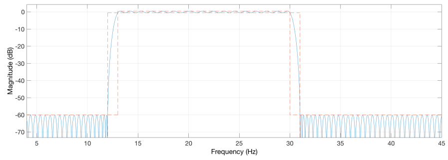
Following the approach of several authors (Hosseini and Khalilzadeh, 2010; Wang et al., 2014; Valenzi et al., 2014; Candra et al., 2015), before the feature extraction process, the preprocessed EEG signals were segmented into equal-length epochs with non-overlapping windows with different sizes. Six different windows were evaluated: 60, 30, 15, 10, 7.5 and 6 seconds. All features discussed next were computed for each epoch. For the classification task, each epoch was labelled with the target emotion of the video that was playing during that epoch.



(a) Theta band (4-8 Hz)



(b) Alpha band (8-13 Hz)



(c) Beta band (13-30 Hz)

Figure 5.3: Magnitude response of the filters used to extract waves of interest.

5.2 Feature extraction

Four groups of features were extracted from each of the 32 channels of the EEG signals: statistical, power, FD and HOC. These features, summarized in Table 5.1, are detailed in the next sections.

Table 5.1: Extracted Features.

Group of Features	Description
Statistical	mean (μ), standard deviation (σ), mean of absolute values of the first differences (δ), mean of normalised absolute values of the first differences ($\overline{\delta}$), mean of absolute values of the second differences (γ), mean of normalised absolute values of the second differences ($\overline{\gamma}$)
Power	using PSD (Welch's method)
Fractal Dimension	applying Higuchi's algorithm
Higher Order Crossings	sequence of filters using the difference (∇) operator, up to order 10

5.2.1 Statistical features

Six commonly-used statistical descriptors, as detailed in section 4.2.3.1, were extracted: mean(μ), standard deviation (σ), mean of absolute values of the first differences (δ), mean of normalised absolute values of the first differences ($\overline{\delta}$), mean of absolute values of the second differences (γ), and the mean of normalised absolute values of the second differences ($\overline{\gamma}$). These were computed from the original EEG signal in the time domain, as well as from the Alpha, Beta and Theta rhythms.

5.2.2 Power features

In order to extract frequency-domain information, the MATLAB implementation of Welch's method (Welch, 1967) was used to estimate the PSD of the Alpha (8-13 Hz), Beta (13-30 Hz) and Theta (4-8 Hz) frequency bands. A Hamming window was used to obtain eight segments in the signal with 50% overlapping samples. The average power of each band was computed by integrating over the PSD estimate, using the rectangle method and used as feature. The total power of the signal was also considered.

5.2.3 Fractal Dimension

To obtain information about the geometric complexity of the EEG signals, the FD for all channels was extracted by applying Higuchi's algorithm (Higuchi, 1988) with $k = 10$. This algorithm has been shown to outperform other methods, such as box-counting and fractal brownian motion, and is commonly used in EEG emotion recognition (Nakisa et al., 2018; Jenke et al., 2014; Sourina and Liu, 2011).

5.2.4 Higher Order Crossings

As explained in section 4.2.3.1, when a filter is applied to a time series, $Z(t)$, its oscillatory behaviour will change, changing the number of zero crossings, as well. This principle is the basis of HOC features, proposed by Petrantonakis and Hadjileontiadis (2010). Based on the approaches followed by Jenke et al. (2014) and Menezes et al. (2017), a sequence of high-pass filters, $\mathfrak{J}_k$, is applied to the EEG signals:

$$\mathfrak{J}_k\{Z(t)\} = \nabla^{k-1} Z(t) \tag{5.1}$$

where ∇^{k-1} is the iteratively applied backward difference operator[1]. Filters up to $k = 10$ were used and the corresponding HOC sequence was obtained by counting the number of zero crossings of the filtered time-series.

The final feature vector is, therefore, comprised of 2304 features, as illustrated by Figure 5.4.

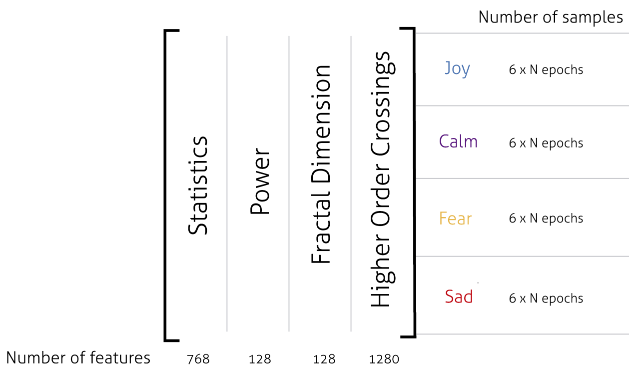


Figure 5.4: Schematic representation of the features matrix with the corresponding number of features and samples for each class.

To avoid biased decisions due to the different magnitudes and units, all features were standardised by calculating the z-scores:

$$z = \frac{x - \mu}{\sigma} \tag{5.2}$$

[1] $\nabla \equiv Z(t) - Z(t-1)$

which produces a zero-mean and unitary standard deviation distribution of the data.

5.3 Feature selection

The substantial number of features resulted from the feature extraction stage calls for a feature selection method capable of reducing the effects of high-dimensional data in the next steps. Feature selection methods are commonly divided into *filter* and *wrapper* methods. Filter methods evaluate the relevance of features by taking into account only intrinsic properties of the data, ignoring the interaction with the classifier, where *wrappers* select features based on the performance of the underlying classifier model. Usually, *filters* require less computational power than *wrappers* and are, therefore, more convenient for big data sets (Jenke et al., 2014; Urbanowicz et al., 2017).

5.3.1 ReliefF

One of the feature selection methods used in this context is the ReliefF. It is a computationally efficient filter feature selection algorithm that is particularly sensitive to feature dependencies and interactions. It is part of a larger family of Relief-based algorithms, whose key idea is to assign a weight to each feature according to the values of k neighbouring samples. The algorithm penalizes the features that give distant values to neighbours of the same class, and rewards those that give different values to neighbours of different classes. Although it doesn't search specifically through combinations of features, this concept of nearest neighbours indirectly accounts for feature interactions (Robnik-Šikonja and Kononenko, 2003; Urbanowicz et al., 2017). The weights assigned to each feature can be ranked and used to select the most important features, thus eliminating features with low discriminative power. This algorithm was explored by Jenke et al. (2014) and Zhang et al. (2016) and, contrarily to multivariate methods, the ReliefF is faster and scalable.

In this work the MATLAB implementation of this algorithm with a value of k set to 10 was used to rank and select the best N features. This value is pointed by Urbanowicz et al. (2017) as the most commonly used.

5.4 Classification

The two most frequently used classifiers were compared for the emotion recognition task, namely SVM and k-NN. In both cases the MATLAB implementation was used. Since this is a multiclass problem, an error-correcting output codes (ECOC) classifier was used. This type of models reduces the problem of classification with three or more classes to a set of binary classifiers, in a one-vs-one approach (Allwein et al., 2000; Dietterich and Bakiri, 1994).

A grid-search was applied to select best hyperparameters for each classifier, according to the mean classification accuracy over all subjects. For the k-NN k parameter, the grid search ranged from 1 to 11 in increments of 1. Regarding the SVM classifier, after some preliminary tests with a linear kernel, which always led to worse results, two other different kernels were systematically compared: a polynomial kernel with polynomial order ranging from 1 to 6; and a radial basis function (RBF) kernel where the kernel scale, *gamma*, is automatically determined by an heuristic procedure in the implementation used. These are the three most used kernels in the literature (Alarcao and Fonseca, 2017). In both searches, the penalty factor, c, was from 1 to 9.

In order to evaluate classification performance when identifying emotional states, the accuracy in a 5-fold cross-validation scheme was computed. This approach randomly splits the training data into five equal-size subsets. In each iteration, four of these subsets are used for training and the other one is used for testing. The process is repeated five times, each time with a different testing subset. This approach ensures that the classifier is not tested with samples that were used for training and also decreases the deviations in the results due to a possibly biased separation of the data. The predictions resulting from the five folds are aggregated and the accuracy is computed as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5.3}$$

where TP, TN, FP and FN denote True Positives, True Negatives, False Positives and False Negatives, respectively. The accuracy can be interpreted as the fraction of predictions that the model classified correctly.

5.5 Results

This section presents the results obtained in the DEAP dataset following the methods previously described. The results represent a subject-dependent analysis, where a new model is trained for each user and then tested using new data from the same subject. Afterwards, a final average across subjects is reported. Following a trend in literature, in the remainder of this work, all reported accuracy values refer to the 5-fold cross-validation accuracy, since this represents a less biased evaluation.

5.5.1 Hyperparameters

In machine learning algorithms and classification systems there is a myriad of parameters that can be tuned to improve the overall behaviour of the system. The hyperparameters of the classifiers tested in this work have already been described in section 5.4. Under the findings of the grid search approach detailed in that section, the model with the highest average cross validation accuracy, a RBF-SVM with $c = 7$, was used for the remainder of this work.

5.5.1.1 Number of selected features

Having set the classifier hyperparameters, the number of features to be selected by the ReliefF algorithm had to be defined. Initially, the mean classification accuracy was computed for the N best features, with N ranging from 10 to 490 in increments of 20. The results are plotted in Figure 5.5. The k-NN (k=3) accuracy was also considered for comparison between the two classifiers.

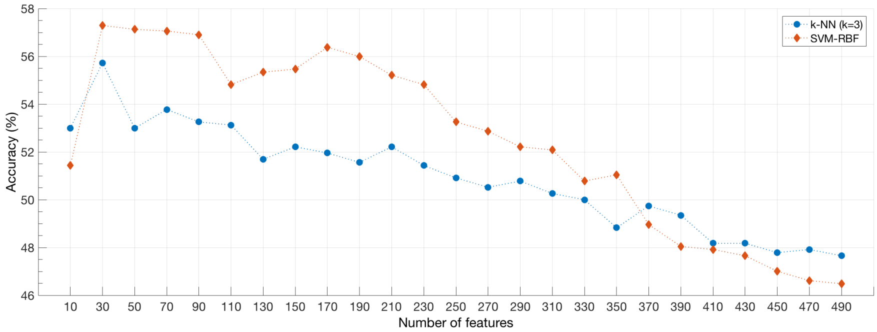


Figure 5.5: Selection of the best number of features in increments of 20.

After a slight increase, both classifiers show a significant decrease in performance as the number of selected features increases. This demonstrates the result of the so-called *Curse of Dimensionality*, a well-known problem in machine learning. For a fixed number of samples, as the dimensionality increases, the volume of the feature space increases exponentially, which makes the samples sparse and an accurate estimation of the classifier's parameters is more difficult.

Seemingly, the highest accuracy was reached when the best 30 features were used. However, this result could be misleading due to the increments of 20 features. Therefore, a new analysis was conducted, this time with N ranging from 10 to 50 in steps of 5. These results are shown in Figure 5.6.

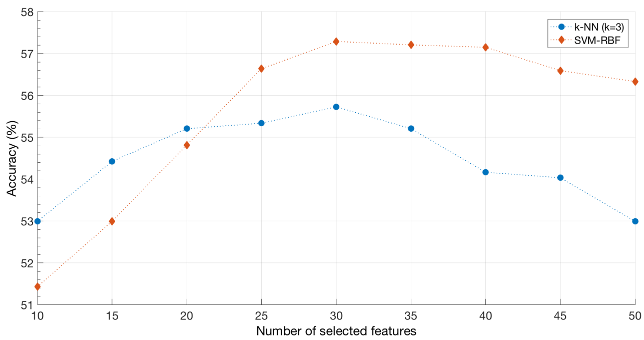


Figure 5.6: Selection of the best number of features in increments of 5.

The second plot confirms that for both classifiers the optimal number of features is 30. Therefore, in the remainder of this work the top-30 features from the ReliefF ranking will be considered. Interestingly, for a small number of up to 20 features, the k-NN has better performance than the SVM. However, this behaviour changes for more than 25 features, where the SVM begins to cope better with this problem and presents overall better performance than the k-NN.

5.5.1.2 Window size

As previously mentioned, before the feature extraction process the preprocessed EEG signals were segmented into various window sizes. Having set the number of features to be used, the classification performance was evaluated with features extracted from these different windowed signals. Table 5.2 summarizes the results.

Table 5.2: Mean classification accuracy for the different window sizes.

Window Size	Mean Accuracy (SD) (%)
60 sec	57,29 (10,84)
30 sec	55,99 (11,27)
15 sec	**78,71 (14,05)**
10 sec	74,61 (12,50)
7,5 sec	78,21 (10,42)
6 sec	77,51 (10,21)

When the entire signal is considered, the average accuracy across all subjects is 57,29%. However, when the signal is segmented in smaller parts, we observe a general increase in performance.

This suggests that relevant information about emotions might be encoded in short time segments, as pointed out by Candra et al. (2015). The authors reported that the effective window size for arousal and valence recognition using the DEAP dataset was between 3–10 and 3–12 seconds, respectively. Within these values, they were able to classify high and low arousal with 65.33% accuracy, and valence with 65,13%. Despite the different methodolgy, the results in Table 5.2 confirm that good results are obtained within this range, particularly for a window of 7.5 seconds, which resulted in an accuracy of 78,21%. These findings can be explained by the fact that a larger window size is more likely to capture more noise and irrelevant data, both of which will obscure any distinct emotional response. Moreover, a large window is more prone to include more than one type of emotion, since the music film clips used in the DEAP dataset evolve over time and the same clip can evoke different types of responses in different moments. Notwithstanding, in this work the best performance is obtained for 15 second segments. With this window size, the maximum accuracy of 78,71% was reached. Despite the minimal difference for the performance obtained with 7,5 seconds windows, a window size of 15 seconds will be used henceforth.

5.5.2 Feature Selection

To evaluate the behaviour of the different groups of features, as well as the effect of feature selection, the system was tested separately for each group of features, without the feature selection step. Table 5.3 summarises the results obtained. When considering only the statistical features, the average accuracy across all the subjects was 55,11%. All other feature groups yielded lower values when tested independently. The use of HOC features resulted in the worst performance of just 39,65%. When all features are used the average accuracy is 53,51%.

None of the feature groups presents a similar performance to that achieved when using the top-30 features as ranked by ReliefF. These results are partially explained by the number of features in each group. For example, the HOC feature group represents 1280 features, an overwhelmingly high number of dimensions which surely impairs the task of the classifier. The results in Table 5.3

Table 5.3: Mean classification accuracy for different sets of features.

Features	**Mean Accuracy (SD) (%)**
Statistics	55,11 (9,43)
Power	44,82 (8,60)
FD	40,69 (7,79)
HOC	39,65 (6,07)
All	53,51 (7,78)
ReliefF (30)	78,71 (14,05)

corroborate the importance of a feature selection stage, not only to reduce the dimensionality of the problem, but also to automatically find the features conveying the most useful information for the task.

In the light of these results, and since there is an increasing appeal for wearable devices with less channels, it would also be interesting to investigate which channels are most frequently selected by the ReliefF algorithm, when ranking the features of each subject. Figure 5.7 presents the map of the 32 EEG channels coloured according to the frequency of appearance in the 30 best features for all subjects.

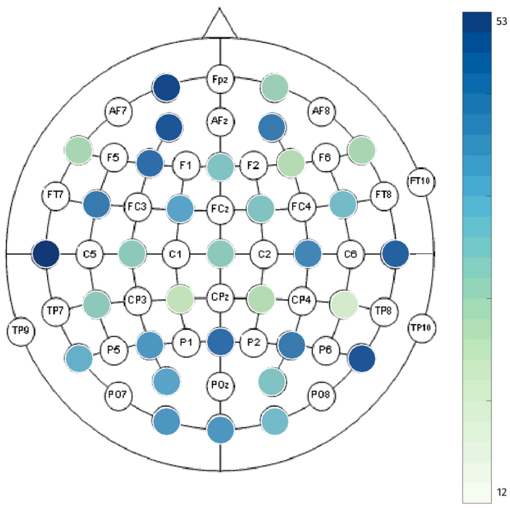


Figure 5.7: Most frequently selected channels. The colour scale represents the frequency of selection, i.e, the amount of times a feature from that channel was selected as one of the best 30, for all subjects.

It is important to note that these channels are not necessarily associated with a specific emotion, but rather indicate an electrode in which one or more features were particularly useful in discriminating between the different classes.

The features from the temporal electrodes T7 and T8 were frequently selected as one of the top-30 features. This finding, also present in the work of Zhang et al. (2016) and Li et al. (2018), is probably related with the type of stimuli used in the DEAP study. The auditory and visual stimuli from the music videos are mainly processed in the temporal lobe.

The sensors located in the parietal area all have a considerably high importance, with Pz, P3, P4, P7 and P8 all falling in the upper end of the scale. This is particularly interesting in the light of the conclusions drawn by both Baumgartner et al. (2006) and Kober et al. (2012), already presented in Section 4.4, who found increased brain activity in the parietal areas of subject who watched immersive videos on screens. This result is also in line with Zhang et al. (2016) who, using the same method for feature selection, reported that most of the top channels were located at frontal and parietal lobes. Other authors using different approaches also mention important findings in the parietal area (Jenke et al., 2014; Lee and Hsieh, 2014; Wang et al., 2014).

The features from the occipital electrodes were also found to be relevant for the discrimination of the different emotions. The location of the visual cortex in this lobe may explain these results, given the visual stimuli presented. Although one could expect a higher activation in the occipital lobe, it is interesting to note that the features from those electrodes, combined with those from the parietal and frontal lobes, are actually relevant to distinguish between the four emotions, to some extent.

It has been shown that the left hemisphere seems to be related to the processing of positive emotions while the right hemisphere is more involved in the negative emotions (Davidson, 2004; Horlings et al., 2008; Liu and Sourina, 2013). However, the electrodes from the left hemisphere seemed to be more informative, specially in the anterior half of the head. This finding is in agreement with the results in Li et al. (2018), who also reported that the left hemisphere generally had a better performance than the right and that the information from the anterior regions enhanced recognition performance.

5.5.3 Emotional state classification

Although it is the most commonly - and frequently the only - reported evaluation metric (Alarcao and Fonseca, 2017), the final average accuracy gives limited insights about the behaviour of the emotion recognition system. Therefore, a more detailed analysis was conducted to explore the results. The plot in Figure 5.8 shows the accuracy obtained for each subject. The dotted line represents the mean accuracy of 78,71% and the solid line indicates the probability of random choice (25%).

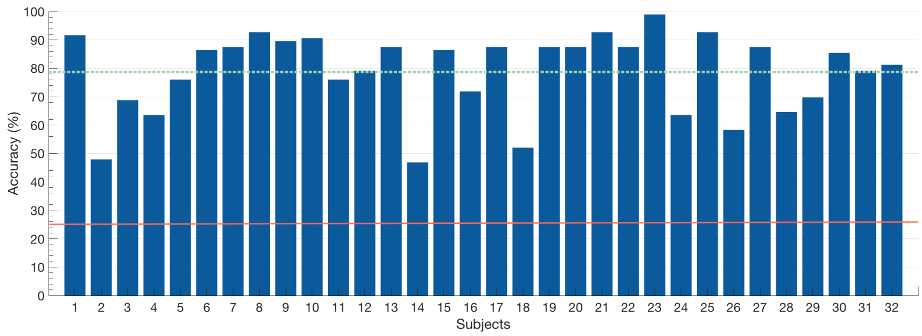


Figure 5.8: Accuracy results for each subject. The green dotted line represents the mean accuracy and the orange solid line indicates the probability of random choice (25%).

In most subjects, the model performs quite close or even better than the average. In fact, in 18 out of 32 subjects the recognition accuracy was above 80%. However, for a few of them the results are undoubtedly worse, barely reaching the 50% of accuracy. These differences can be, to some extent, associated to the individual characteristics in the EEG signals, motivated by anatomical and neurophysiological properties. Despite the clear differences in performance for the different subjects, it is important to note that in all of them the model was able to largely exceed the probability of random choice which would be 25% for the 4 classes considered.

For a more intuitive understanding of the results, the confusion matrices for each subject's classification were computed. A graphical representation is presented in Figure 5.9, where the values are encoded in a colour map. In each matrix, the vertical axis corresponds to the true class and the horizontal axis to the predicted class (Joy (J), Calm (C), Sad (S) and Fear (F)). For direct visual comparison the ideal classification case, where all samples are correctly assigned to their class, is presented in the bottom-right corner. A quick glance over the matrices confirms the contrasts in performance for the various subjects. Most of them show very similar patterns to that of the ideal case, but there are some in which the number of misclassified samples is quite significant, such as subjects 2 or 14.

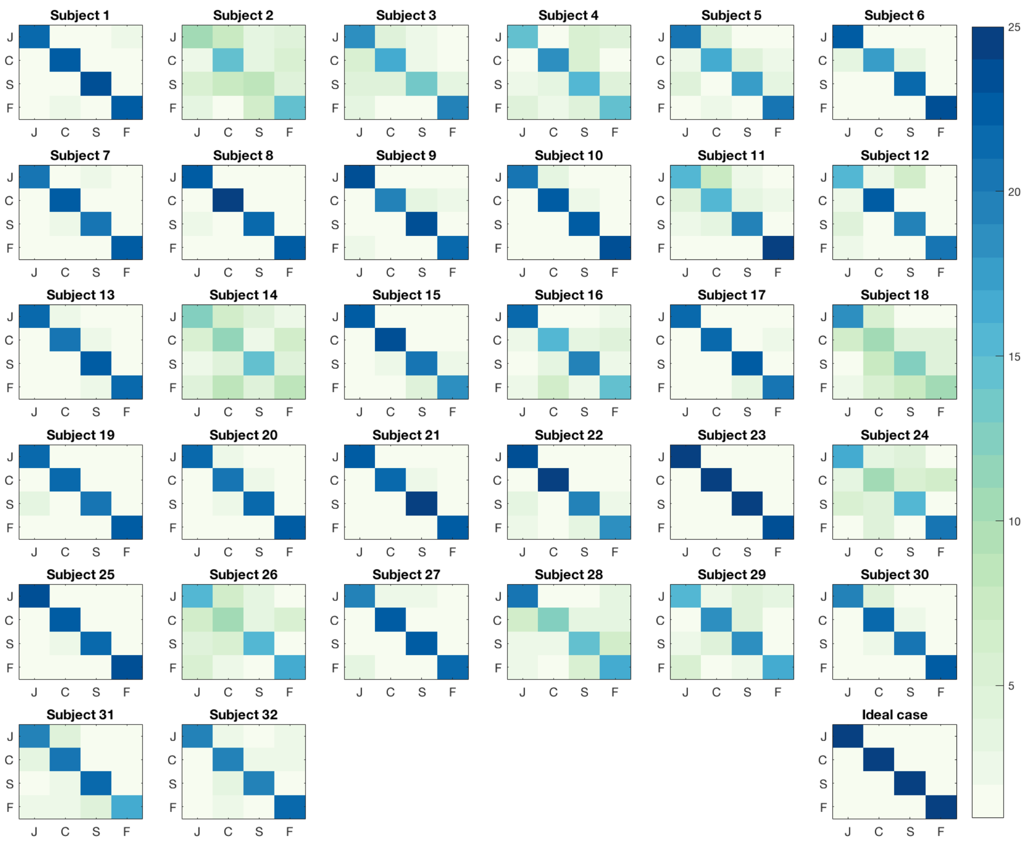


Figure 5.9: Graphic representation of confusion matrices for each subject's classification using the predictions from the SVM Classifier. In each matrix, the vertical axis corresponds to the true class and the horizontal axis to the predicted class: Joy (J), Calm (C), Sad (S) and Fear (F).

A confusion matrix makes it easier to evaluate the differences in performance for each class and to compute a series of metrics that give a more complete picture of the model's performance. Therefore, for each class, the accuracy, precision, recall and F1-score were computed in each subject following Equations 5.3 to 5.6. Table 5.4 displays the results, averaged across all subjects.

$$Precision = \frac{TP}{TP+FP} \tag{5.4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5.5}$$

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{5.6}$$

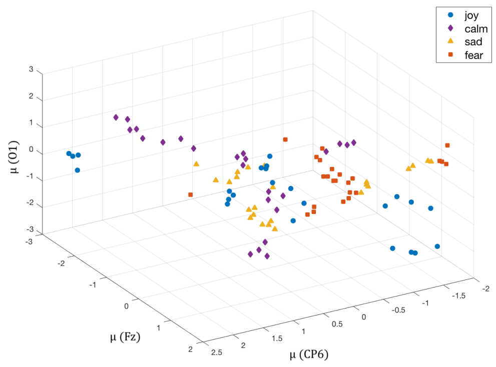
Although very similar, precision and recall represent different interpretations of the performance of the system. Precision refers to the proportion of positive identifications that was actually correct. On the other hand, recall represents the proportion of positives that was identified correctly. Looking at the example of class *Fear*, the results indicate that when the model predicts that a sample is from the class *Fear*, it is correct 81,94% of the times. Nonetheless, it correctly identifies 79,82% of all *Fear* situations.

Table 5.4: Performance metrics evaluated by class.

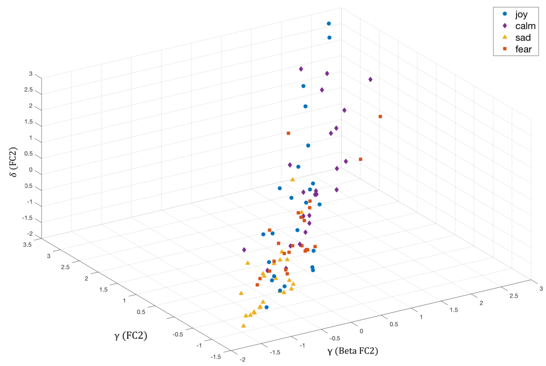
Class	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Joy	89,62	79,47	79,56	79,51
Calm	88,05	76,26	76,95	76,60
Sad	89,16	78,15	78,52	78,33
Fear	90,59	81,94	79,82	80,87

Usually, there is a trade-off in improving precision and recall. That is, improving one typically reduces the other. This means that both metrics have to be considered for a correct interpretation of the results. In order to simplify this analysis, the F1-score is computed. It represents the harmonic mean of precision and recall, meaning that it indicates the balance between the two. The results in Table 5.4 confirm that the model is balanced and achieves overall satisfactory performance. Nevertheless, the *Calm* class is the one with the worst results in all metrics, indicating this was the harder emotion to discriminate. Looking at the confusion matrices, this emotion is mostly mistaken with *Joy* and *Sad*. These are the most similar emotions to the *Calm* state since, according to the valence-arousal plot, they are both adjacent to the *Calm* quadrant. Even intuitively, it is more difficult to distinguish between these emotions when describing feelings.

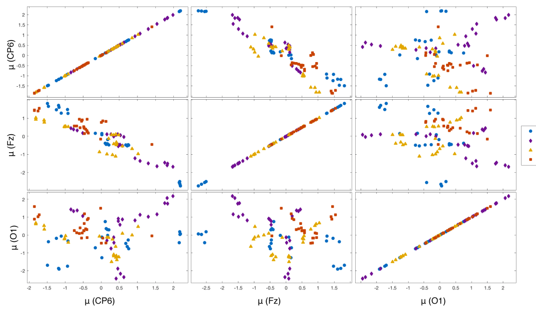
Despite the good overall accuracy, the individual performance indicates a significant degree of variation between subjects. Exploring the results from subject 23 and 14, those with the best and worst result, respectively, could partially explain these differences in performance. Figure 5.10 depicts 3D and 2D scatter plots of the samples represented along the three best features selected for each subject. For subject 23 these were the mean value of the signals for electrodes CP6, Fz and O1. The three best features of subject 14 were the γ of the Beta waves in channel FC2, and the γ and the δ of the unfiltered signals in the same channel.



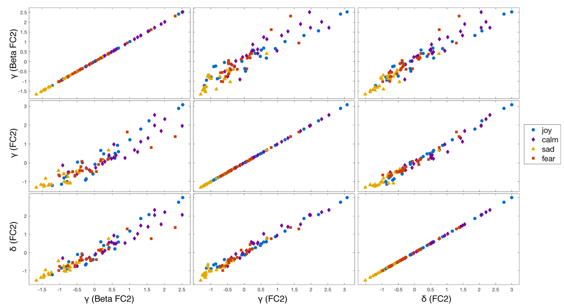
(a) 3D scatter plot for the three best features selected for Subject 23.



(b) 3D scatter plot for the three best features selected for Subject 14.



(c) Corresponding 2D plots for Subject 23.



(d) Corresponding 2D plots for Subject 14.

Figure 5.10: Scatter plot of the three best features for the subjects with the best (a) and worst (b) results. For fair comparison, in both cases the plots represent the three best features selected for the subject.

By observing the spatial distribution of the samples of each class, one can intuitively understand the difficulties in the classification task. Considering that the classifier is using a one-vs-one approach, it is fairly easy to imagine a separation boundary between the classes in subject 23. However, that is not the case for subject 14, whose classes are not minimally separated. This obviously undermines the behaviour of the classifier, especially taking into account that these are allegedly the features with the best separation between classes, meaning that in all other dimensions the case is even worse. Furthermore, the features selected for subject 14 are noticeably correlated. This is a problem for the classification task, since the information is redundant. It also reveals a known drawback of the ReliefF algorithm which selects features regardless of whether some are strongly correlated with others (Urbanowicz et al., 2017).

In a wider analysis, the inter-subject variability of the results can also be explained by the intrinsically subjective nature of the task. To begin with, not everyone has the same emotional response to a common stimuli, as it is associated with past experiences and memories (De Kloet et al., 2005). Also, the self-assessment scale used to gather affective ratings is subjective and prone to different interpretations based on the impressions of the participant. It is often the case that people have difficulties in articulating their actual emotions and associated states. For example, one can be inclined to rate something as negative in terms of valence due to social stigma, when, in fact, it is not negatively perceived. Therefore, it could be that some of the subjects could not precisely decipher their actual emotional state and translate it to the SAM scale. Furthermore, the classification labels are based on the mean ratings across participants in the study and not to the subject-specific rating, which can also explain the bad results for some of them. For instance, what was labelled as *Joy* could have actually been perceived and rated as *Calm* for a particular subject.

5.5.4 Subject-independent classification

The good results obtained in a subject-dependent analysis can be overly-optimistic. For any valuable future application of an emotion recognition system, the ideal is not to have to rely on any kind of subject-specific calibration and hence, a subject-independent model is of great importance. Table 5.5 presents the results of the aforementioned pipeline - with the same hyperparameters - when applied to all subjects. In this approach, here called "Global", the classifier is fed with the samples from all subjects and then the same 5-fold cross validation scheme was applied for evaluation.

Table 5.5: Performance in different classification schemes.

Scheme	**Accuracy SVM(%)**
Average across subjects	78,71
Global classification	54,98

The accuracy of 54,98%, although still considerably greater than the probability of random choice, reveals a substantial decrease in performance when compared to the 78,71% for the subject-dependent analysis. However, it must be noted that these results serve merely as a comparison, since the parameters were probably not the optimal for this type of approach as they were for the subject-dependent results. Yet, because of the inter-subject variability already discussed and the amount of possible variations of the system, it would still be difficult to optimize global parameters that can be effective. The problems presented above, such as the subjective and possibly erroneous characteristics of the ratings, are also present in this situation and have repercussions in classification performance. Despite these factors, the results are in line with those reported in the literature. Li et al. (2018) achieved a similar recognition accuracy when considering a "leave-one-subject-out" strategy: 59,06%. Very few authors have compared both subject-dependent and subject-independent. The work of Sohaib et al. (2013) reports a performance of 56,10% for recognition of the same four emotions across 15 subjects in their own experimental data. When considering each subject independently, their best results were obtained with a k-NN classifier, which yielded 70,20% accuracy. However, this represents the average of only 3 subjects. Other works, such as the ones by Lin et al. (2014) and Jie et al. (2014) also show similar trends, highlighting the difficulty in developing a *universal*, subject-independent system.

5.6 Final remarks

In this chapter, a subject-dependent emotion recognition model was proposed. The results reveal a good general performance, given that the algorithm can recognise 4 emotions with 78,71% of accuracy. The results of some of the most relevant works in the literature are illustrated in Figure 5.11. It is difficult to make direct comparisons among classification studies in emotion recognition since there is a great degree of variability, specially in the data acquisition protocol, but also in the approaches chosen for classification. However, for the most similar works, the results are better than those reported in the literature. The highest mean accuracy of 78,71% is better than the results obtained by Mehmood and Lee (2016) who tried to recognise the same set of emotions. Nakisa et al. (2018) also tested their approach in the DEAP dataset, but with inferior performance that the one presented in this chapter. Other works have reported similar results. Liu and Sourina (2016) tried to apply their algorithm for a wider range of emotions. Yet, for the 4-class problem, they report an average accuracy of 80% on the DEAP dataset, which is fairly close to the result shown in this work. Valenzi et al. (2014) reports excellent results. However, these were not obtained in the DEAP data and only 9 participants were involved.

The evaluation of results for a subject-independent situation exposes the difficulties in developing a “universal” emotion recognition system from EEG signals. The variability found across subjects impairs the optimization of global parameters that can be effective. However, the performance of the proposed algorithm in this case, even with suboptimal parameters, was 54,98%. This is very similar to the result obtained in the work of Li et al. (2018) for the same dataset and is in line with the few works which have compared both subject-dependent and independent approaches (Sohaib et al., 2013; Lin et al., 2014; Jie et al., 2014). Also, the results are influenced by the labelling scheme which can be biased for some subjects, since the mean ratings do not necessarily represent the evaluation give by a particular person. To address this limitation, the use of unsupervised learning algorithms for classification could be investigated.

All in all, despite the limitations of such an intrinsically personal and subjective topic as are emotions, the methodology and results presented in this chapter validate the possibility of exploring physiological signals to identify psychological states and show encouraging results.

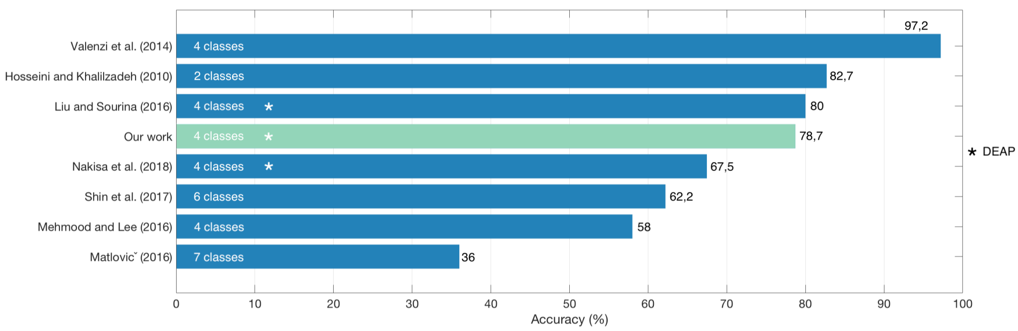


Figure 5.11: Comparison with literature results. The works which report their results in the DEAP dataset are marked with an asterisk (*). The number of classes (emotions) considered is also indicated.

Chapter 6

Visual Stimuli Modalities in EEG Emotion Recognition

In order to assess the brain response to different modalities of visual stimuli (images, 2D videos and spherical VR videos), a study was designed so as to record the EEG signals of participants who were presented to those stimuli. Based on the valence-arousal model, represented in Figure 1.1, the target emotions were the same as those studied in the previous chapter: Joy (HVHA), Calmness (HVLA), Fear (LVHA) and Sadness (LVLA). This chapter details the protocol used for emotional stimulation and wearable EEG recording and presents the results obtained.

6.1 Participants

A total of 17 healthy volunteers, with a mean age of 25 (SD=8), participated in this study. From these, 11 were females and 6 were males. Due to the limited time frame for the development of this thesis, it was only possible to record the EEG from 8 subjects.

6.2 Stimulation protocol

The stimulation protocol consists of three main phases, illustrated in figure 6.1, each with a different visual stimuli modality: images, 2D videos and spherical VR videos.

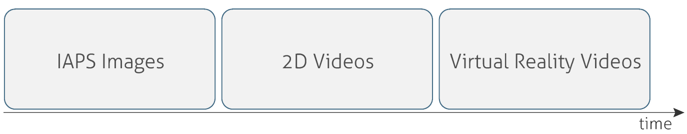


Figure 6.1: Main phases of the stimulation protocol.

The participants sat comfortably in a swivel chair and faced a computer screen, placed in a quiet room with appropriate lighting conditions. The subject is first presented to still images, chosen from the IAPS database in order to evoke specific emotional responses. Afterwards, a set of four videos is displayed, each corresponding to one of the emotional states. Finally, the subject visualizes another four videos, this time in a VR environment, wearing the Samsung Gear VR headset. For 8 of the participants, during the experiment the EEG was recorded with the Emotiv Epoc+. A picture of a participant in two different phases of the experiment is shown in Figure 6.2.

The same person does not see the same video twice, since it would hinder the emotional effect. For each phase, the participants also filled self assessment questionnaires, using the 9-point SAM, identifying the valence and arousal levels they experienced during the experiment. To guarantee that the participants would indicate meaningful ratings, the SAM rating scale was carefully explained beforehand.

The first two parts of the protocol were designed in OpenSesame, an open-source program developed to create experiments for psychology, neuroscience, and experimental economics, with a graphical user interface and Python scripting capabilities (Mathôt et al., 2012).

Figure 6.2: Participant during the experiment. The Gear VR headset is used for visualizing videos in a VR environment. The EEG data is being recorded with the Emotiv Epoc+.

6.2.1 IAPS Images

The IAPS, briefly mentioned in Section 3.1, is a large set of 192 photographs selected to provoke emotional responses for studies on emotion and mood psychology. Each picture is associated with the corresponding mean ratings for valence and arousal, according to the evaluation by the participants in the study by Lang et al. (2008). In this work, a total of 80 images were chosen to represent the four emotions (joy, calmness, sadness and fear) based on their average ratings. The ratings of the selected images are plotted in Figure 6.3. These images were chosen so as to be as equivalent as possible to the videos used in the next sections. Images with ratings very close to the central point (Valence = 5, Arousal = 5) of the plot were not included, and repetition of the same theme was avoided. Also, some specific picture groups, such as mutilations and erotic content, were excluded. As an example, Figure 6.3 also presents similar images to those used.

The IAPS images were shown to the participants following the protocol illustrated in Figure 6.4. After some initial instructions where the task is explained and the participant is asked to avoid

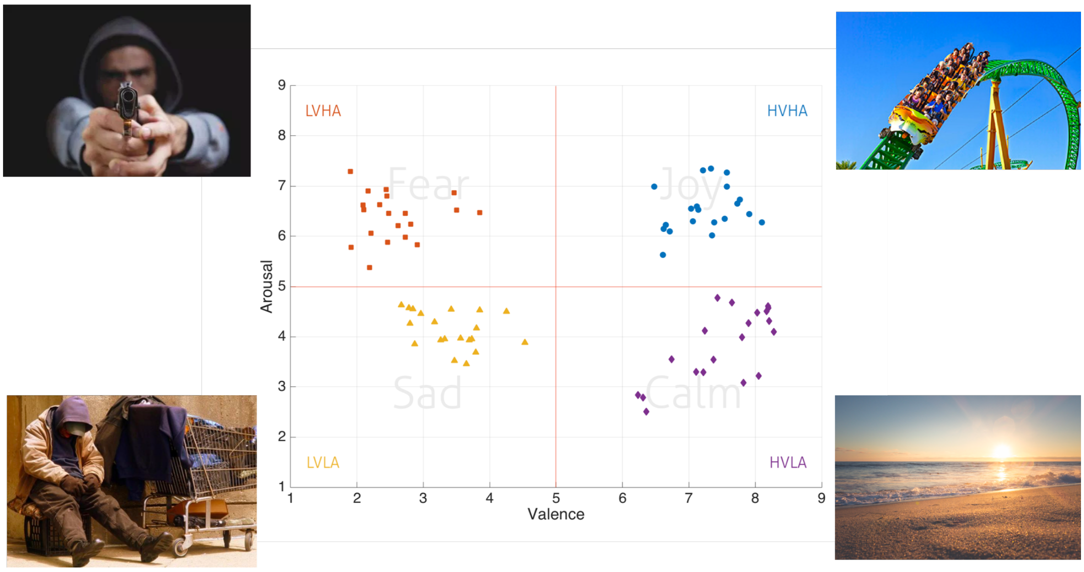


Figure 6.3: Ratings of the chosen IAPS images and corresponding examples.

movement and blinking, two short *beep* sounds are played with an interval of 5 seconds. The participant is asked to close the eyes after the first sound and open just after hearing the second. This contributes to draw the attention of the patient, while also working as a synchronization marker and quality check, since the characteristic Alpha waves will be easily distinguished in the EEG signal. The participant is then asked to observe an example neutral image and rate it using the 9-point SAM scale in Figure 6.5 so as to understand how to use the rating form.

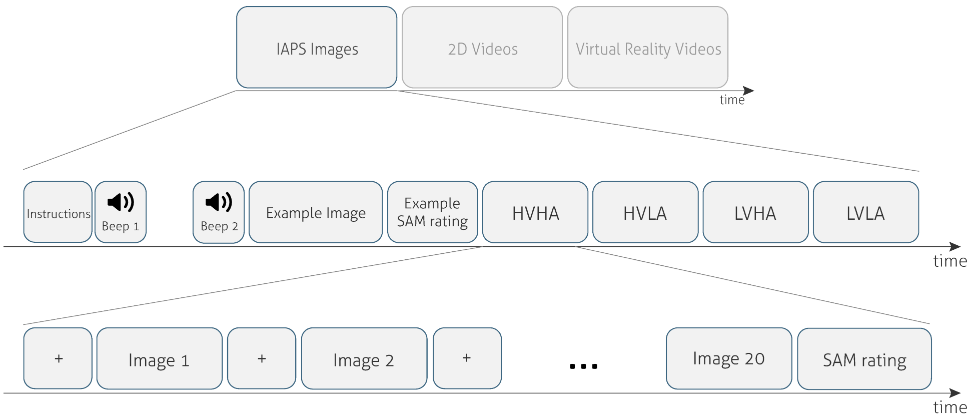


Figure 6.4: Detailed protocol: IAPS images.

After completing this task, the 20 IAPS images belonging to the same emotion group are randomly presented for 6 seconds, accordingly to the original IAPS protocol (Lang et al., 2008), making it enough to elicit the emotional reaction, but not too long so as to produce habituation to the image. Between each image, a fixation cross was shown for 1 second. After 20 images, the participant is asked to rate the overall feeling. This is repeated for each of the four emotions.

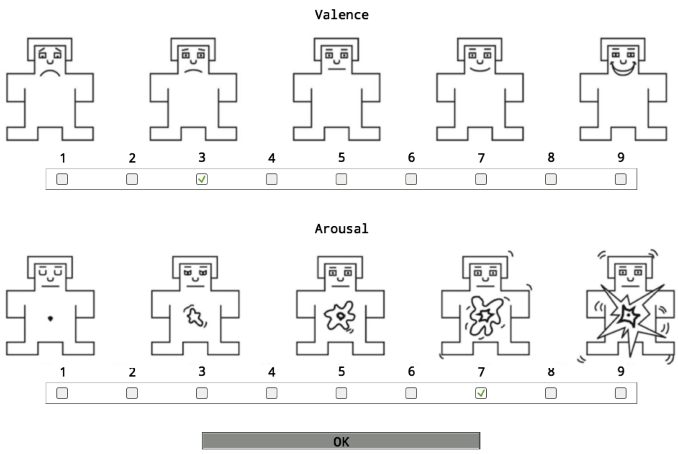


Figure 6.5: Form designed for evaluation with the SAM rating scale.

6.2.2 Videos

The recent work of Li et al. (2017) establishes a public database of immersive VR video clips and corresponding valence/arousal ratings that can be used for studies on emotion recognition using virtual reality. In the proposed set of videos, only two fall in the low-valence, high-arousal quadrant, meaning that only these two can be used to provoke fear. For each of the other emotions, three videos were chosen according to their valence and arousal ratings similarly to the IAPS images, making a total of 11 videos as shown in Figure 6.6. These 360º VR videos were converted to the normal flat format, so as to allow for the comparison between a normal visualisation mode and a virtual environment.

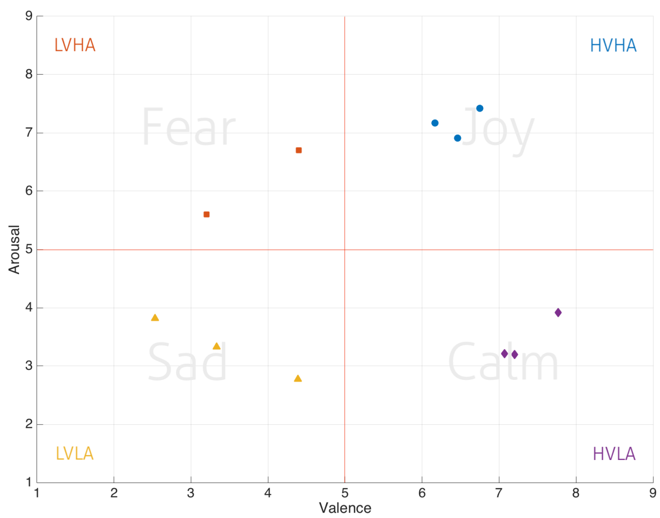


Figure 6.6: Ratings of the chosen videos.

For the 2D videos stimulation, the participants followed the protocol depicted in Figure 6.7. Similarly to the images protocol, after the initial instructions and *beep* sounds, the experiment starts with a neutral example. Afterwards, one video from each emotion group is randomly selected and shown to the participant. When the video is finished, the participant rates it using the same SAM rating form mentioned before.

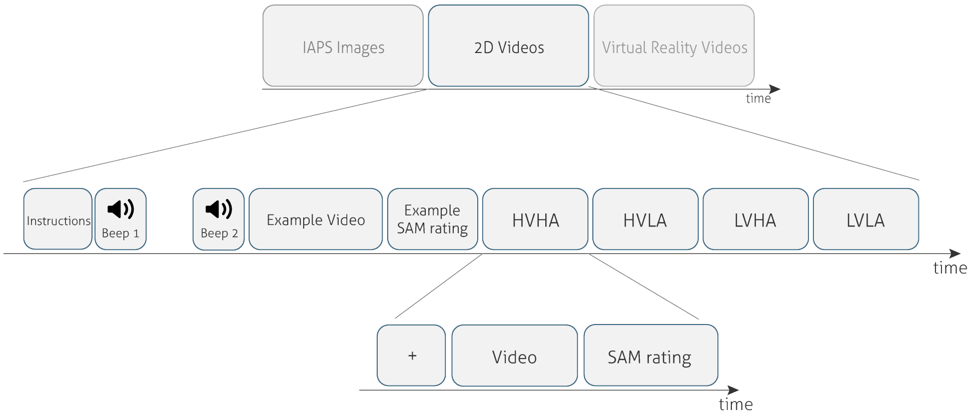


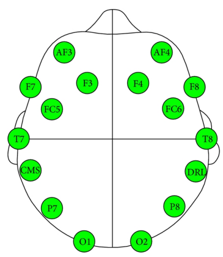
Figure 6.7: Detailed protocol: 2D Videos.

The stimulation with the 360° VR videos followed the same principle, although with a less automated protocol since the video had to be selected directly in the smartphone connected to the Gear VR headset. This was done ensuring that the video was different to the one previously seen for that target emotion and that the different videos were seen by approximately the same number of participants. In this phase of the protocol, the swivel chair allowed the participants to turn around to explore the virtual environment with minimal effort and head movement, thus reducing the noise and movement artifacts in the EEG signal as much as possible.

6.3 EEG recording

During the stimulation protocol, the EEG is recorded at 128 Hz using the Emotiv EPOC+ headset and its Testbench software. This device is designed for contextualized research and advanced brain computer interface (BCI) applications. Shown in Figure 6.8a, it is an EEG recording system with 16 gold-plated contact-sensors that are fixed to flexible plastic arms and positioned according to the international 10-20 system, as represented in Figure 6.8b (EMOTIV, 2018).

(a) Emotiv EPOC+ headset.



(b) Electrode placement according to the international 10-20 system.

Figure 6.8: Emotiv EPOC+ headset and electrode placement.

Using the Python scripting from OpenSesame, a set of triggers marking the start and end of each picture or video was sent to the EEG recording software (Testbench). This was done so that the markers could be used for identifying the regions of interest in the whole recording. For this purpose, two virtual serial ports were first configured, creating a communication channel between OpenSesame (which sends the trigger) and Testbench (receives the trigger).

6.4 Emotion recognition

Using the same pipeline described in the previous Chapter (applied to the DEAP dataset) we studied the differences in emotional recognition between the 3 different stimuli modalities (images, 2D videos and VR) and how it compares to the previous results. Due to the differences in the protocols, some modifications had to be included. These are explained in the next paragraphs.

6.4.1 Preprocessing

Contrarily to the DEAP data, which had already been cleaned and filtered, our experimental data had to be preprocessed so as to remove noise and artifacts. This was done using EEGLab (Delorme and Makeig, 2004), analogously to what was described in Section 5.1. Eye movement artifacts were manually removed, a 4.0 - 45.0 Hz bandpass filter was applied and the data was averaged to the common reference. The Alpha, Beta and Theta bands were extracted using the same bandpass filters.

6.4.2 Samples and epochs

In the first part of the protocol the participants watched a total of 80 images for 6 seconds each. Therefore, instead of considering the 15-second windows as before, the features are extracted from the 6-second EEG data corresponding to each image and total of 80 samples are fed to the classifier (20 for each class). This is approximately the same number of samples per class used in the DEAP dataset analysis (24).

Regarding the 2D and VR videos, in order to have comparable results to those obtained in the DEAP study, the window size of 15 seconds was maintained in this analysis. For most of the videos, at least six 15-second segments could be extracted. However, one of the drawbacks of the videos proposed by Li et al. (2017) is their significant differences in duration. This meant that the windowed epochs had to be adjusted for the shorter videos. In those cases, the overlapping of the windows was adjusted so that six epochs could be present.

All the other parameters mentioned in the previous Chapter were the same, namely the extracted features, the number features to be kept after the ReliefF ranking and the SVM hyperparameters.

6.5 Results

This section describes the results achieved. In this part of the work the fundamental objectives were to assess the performance of the methodology devised earlier in a new, distinct dataset and to investigate how would the emotional response vary with three different modalities of visual stimuli.

6.5.1 Emotion recognition with different visual stimuli

As explained before, the newly acquired data was used to evaluate the performance of the previously devised methodology in a new scenario. Table 6.1 shows the results obtained in the three different modalities. Again, these results represent a subject-dependent approach, with the average across subjects, 8 in this case, being reported.

Table 6.1: Average classification results for different stimuli modalities.

Class	Mean Accuracy (SD) (%)
Images	72,00 (16,19)
Videos (2D)	82,74 (7,04)
Videos (VR)	76,67 (10,76)
DEAP (See Chapter 5)	78,71 (14,05)

Overall, the results are quite similar to the performance achieved in the DEAP dataset in equivalent conditions. This reveals that the algorithm was sufficiently able to deal with the new data, which has less channels and was acquired with a totally different, wearable equipment. However, the standard deviation values still indicate a considerable degree of variation between subjects, as more thoroughly discussed in the previous chapter.

The videos, both in 2D and in VR, produced a better classification performance. It would be tempting to say that this is a result of a “stronger” effect on the participants, which could have led to “clearer” features. Although this explanation can not be completely disregarded, it is most probably the case that this is a result of the overlapping of the windowed segments in the videos. This may have had introduced some bias in the training phase of the classifier, since the samples share similar feature values, explaining the better results.

The VR videos show a decreased accuracy when compared to the 2D ones. Although a fully direct comparison can not be made since the same videos were not seen by the same person in both modalities - and even if they were, the stimulus effect would not be the same - the difference can possibly be explained by some of the problems of the third phase of the protocol. To begin with, the fitting of the VR headset pushes some of the electrodes and may slightly change the position of others. Also, despite the use of the swivel chair, there is a significantly higher movement of the head and body of the participant while exploring the virtual environment. The proximity of the smartphone to the EEG electrodes may also cause some interference in the measurements which are not present in the 2D videos scenario. All these factors introduce noise and artifacts into the EEG recording, which may help explain the worse results if compared to the 2D situation.

6.5.2 Analysis of SAM ratings

The subjects rated their perceived emotions using the 9-point SAM scale. Although the number of subjects in this part of the study (17) is not very representative, it is still quite acceptable to gain some insights about the use of the different visual modalities.

To have a general view on how these ratings compared to the original ones, here treated as *ground truth* (GT), the means of the ratings were plotted in the valence-arousal space in Figure 6.9. The GT corresponds to the mean ratings in both dimensions for the selected images and videos, as presented in Images 6.3 and 6.6.

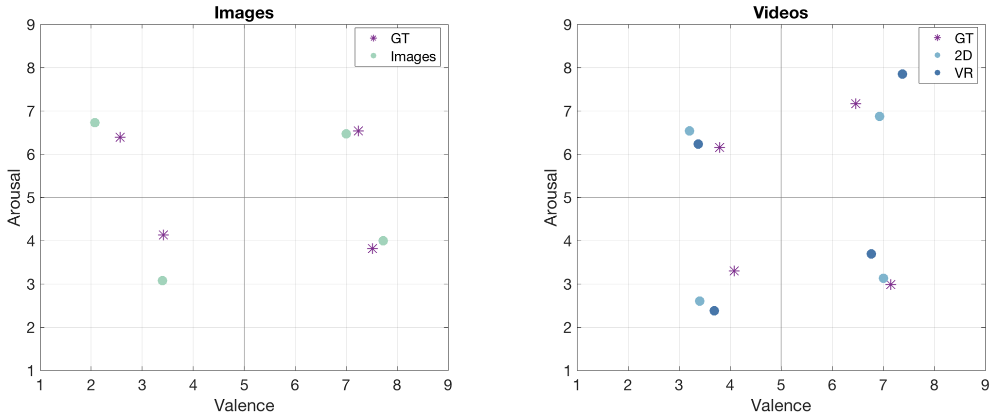


Figure 6.9: Comparison between the original (GT) mean ratings and the ones provided by the participants for the IAPS images (left) and the videos (right).

When considering the IAPS images, the mean ratings for the high-valence emotions are particularly similar to the GT. There is a small offset in valence and a minimal difference in arousal. For the negative-valence side of the plot, the ratings of the participants were slightly more deviated from the original ones. As far as the videos are concerned, the ratings were closer to the GT for the *Fear* and *Calm* situations. Nevertheless, overall the participants' assessment is quite close to what was expected, indicating that the stimuli had the intended effect on the subjects of this study.

In order to understand how each emotion was perceived in the different modalities, the mean valence and arousal ratings were evaluated through unpaired-sample *t-tests* in order to find the statistically significant differences between images, 2D videos and VR conditions. The results were significant for $p < 0.05$ and the results are summarized in the bar graphs in Figure 6.10.

When considering the *Joy* stimulus, although the VR videos resulted in a slightly higher mean value, the mean valence ratings are show no significant differences between the three scenarios. However, the VR videos seem to have induced stronger feelings of arousal in this case. The differences are significant when compared to both the 2D videos and to the images.

For the *Calm* case, there are significant differences between the images and both types of videos. In this case, the images were rated as more positive in valence than the videos, which were

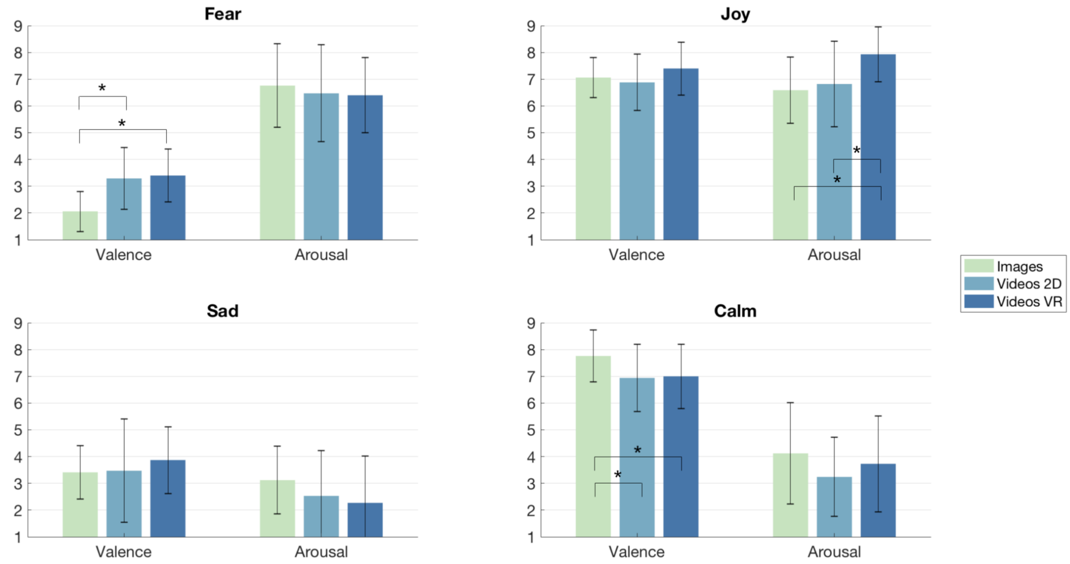


Figure 6.10: Bar graphs (means and standard deviation) comparing the three different modalities. The pairs with significant ($p < 0.05$) differences are indicated.

almost equally perceived. The same happened for the arousal dimension, although the differences were not significant.

The *Sad* valence ratings are almost the same for the three conditions, despite a minor difference in the VR videos which were slightly more neutral (Valence = 5). In the arousal dimension these were the ones with the lowest mean value, followed by the 2D videos and the images. Nevertheless, none of these differences were found to be significant in the *t-tests*.

Finally, for the *Fear* stimuli, the images were significantly more negative than the videos, which were perceived with identical valence in both modalities. The arousal ratings were also quite similar, with a somewhat higher value for the images.

Overall, there is not indisputable evidence that, as could be empirically expected, the videos are better than pictures in inducing emotions. The same conclusion was pointed by Dhaka and Kashyap (2017) and Uhrig et al. (2016) who found that film clips were less effective in producing the corresponding emotional state. One possible reason for this is the length of the movie clips, which can end up fading the desired stimulus. In many of the videos, there is usually a moment of greater emotional intensity, but this is then overthrown by a more neutral conclusion. By the time the participant is asked to rate the video, he has already processed that peak of emotion and tends to indicate a more neutral rating than would be expected. Also, some of the clips are understood as telling fictional stories, which leads to a more distant relationship with the topic.

6.5.3 Analysis of EEG recordings

Although the ratings are not completely clear about whether the different modalities of visual stimuli are perceived in a distinct manner or not, they only represents a discrete, explicit measurement which can be influenced by several factors, as already discussed in Section 5.5.3. The visual stimuli certainly produce more complex reactions than those that can be told in the form of valence and arousal. It is therefore important to investigate more objective and undeceivable measures. For this reason, an exploratory analysis was conducted so as to understand the differences in the EEG signals in response to the three visual stimuli. Following a literature trend, the power for the Alpha, Beta and Theta frequency bands was computed for each modality and for each emotion separately, using the same method applied before and described in Chapter 5. The average values and corresponding standard error bars are shown in Figure 6.11.

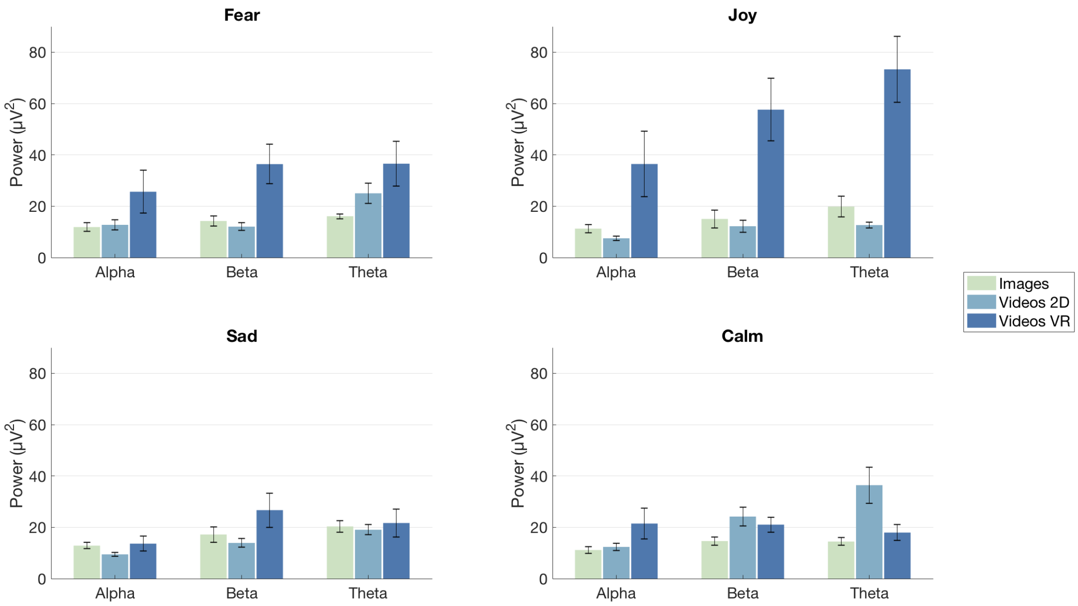


Figure 6.11: Bar graphs (means and standard errors) showing the power for the three frequency bands in each modality. The values are averaged across all the electrodes, participants and stimuli.

A quick glance over the results can reveal that for all the emotions, the images and 2D videos show quite similar values in all the frequency bands. However, for *Fear* and *Joy* in particular, there is a considerable and general increase in power for the VR videos. When looking in more detail to the *Joy* plot, this increase is more pronounced for the Theta band. Slobounov et al. (2015) also found a higher Theta power in the frontal areas of the brain when studying sensory-motor integration in a 3D VR environment as compared to a 2D display. This increase seems to be related with spatial navigation tasks and is associated with the sense of presence (Kahana et al., 1999; Kober et al., 2012). When the participants were immersed in the virtual environment, there was a greater effort and desire to explore the spherical image, specially when the video is more arousing, and this appears to be confirmed by the increased Theta activity in both *Fear* and *Joy* conditions.

Interestingly, the emotions for which there is a bigger difference in power for the VR videos when compared to the other modalities are the ones which are supposed to induce a higher level of arousal. When observing the *Calm* and *Sad* scenarios, the power in the frequency bands is similar across the different stimulus modalities, indicating that in low-arousal situations, there is not a compelling advantage in using VR when compared to normal 2D videos or pictures. However, in high-arousal scenarios, the EEG signals seem to capture the increased excitement.

It should be noted, however, that the variations in the individuals' neurophysiological and anatomical properties led to considerable differences in the absolute power of the EEG frequency bands. This is reflected in the significant variance of the values, as represented by the standard error bars. Also, the reduced population in analysis, as well as the higher proportion of women, may have hindered or masked the results.

6.6 Final remarks

This chapter mainly explored the influence of the emotion-inducing modality on the experience and perception of emotions, by looking at both the SAM ratings provided by the participants and the EEG signal recorded during stimulation. The ratings proved to be similar to the original ones, indicating that the visual stimuli had the desired effect. However, the results for the different modalities did not reflect a pronounced effect on the perception of the different emotions. Although some significant differences were found between images and videos, the results varied across emotions and could not be generalized so as to support a sound conclusion. As discussed, the length of the videos may have influenced the results. Further investigation on this topic would benefit from a more carefully designed experiment that has this factor into account.

On the other hand, the analysis of the average power of the EEG signals revealed a considerable increase when the participants watched VR videos with respect to *Joy* and *Fear*, two emotions which share a high level of arousal. These results suggest that the concept of arousal, which is more difficult to perceive and express, is well captured by the EEG power features. This also points out to the benefits that VR environments can have in all sorts of engaging activities.

Despite the encouraging results, this study presented several limitations. The reduced number of participants undermines the generalization of the results and the higher proportion of female participants may have introduced some bias in the analysis, considering that women tend to be more emotional than men. A study with more participants and a better distribution in both gender and age is needed in order to validate these results. Also, the power features computed to reach these conclusions represent the average across subjects, in which there is high variability, but more importantly, the average across electrodes. Therefore, the underpinnings and details of the response to VR videos couldn't be fully explored.

Chapter 7

Conclusions and Future Work

One of the main goals of this dissertation was to develop a system for psychophysiological state identification, more precisely one that could classify human emotions using brain activity. The results obtained in a publicly available dataset corroborate the feasibility of using EEG signals for this purpose. An accuracy of 78,71% was achieved when considering a subject-dependent classification. Although comparison between studies in the literature is difficult, the proposed methodology can surpass some of the most similar works in the area. In addition, the same algorithm proved to be robust in new data, acquired with a different device and in different conditions, and effective in various stimulation modalities.

The difficulty in developing a "universal" system was highlighted when testing the algorithm in a subject-independent manner. The best accuracy obtained was 54,98% and, while being similar to state-of-the-art results, it still indicates that there is clearly much room for improvement when it comes to emotion recognition systems that can be applied in real scenarios.

The work devised in the dissertation also seeked to assess the potential of using VR environments for emotion-related EEG studies by comparing it to the use of standard pictures and videos. To date, few works have compared such modalities using EEG signals rather than subjective self-assessment questionnaires. The results revealed that when analysing the SAM ratings, the stimuli in all three modalities are very similarly rated. On the contrary, the power of the frequency bands of the EEG signals were significantly different for the VR videos. This suggests that the EEG features can capture emotional responses which can not be expressed in the form of a rating.

All in all, despite the troublesome details of such an intrinsically personal and subjective topic as are emotions, the work developed in this dissertation validates the possibility of exploring physiological signals to derive information about psychophysiological states, which can ultimately lead to a better understanding of our deep affective nature and serve in such diverse applications such as *eHealth*, entertainment, education or marketing. Furthermore, this dissertation also hinted that these signals can reveal facets of such mental states that are not even sufficiently acknowledged or expressed by oneself, judging by the difference in the effect of the VR videos when considering the ratings and the EEG signals.

Besides addressing the limitations discussed throughout this document, future work should explore other physiological signals such as the ECG, which as been widely proven to accurately detect stress. With the massification of VR systems in neuroscience research, the investigation of the perception of different emotional stimuli using other techinques such as fMRI can also be an interesting line of research. All these suggestions could complement the work developed in this thesis and shed some more light on the still unconquered realm of affective neuroscience and its contributions to affective computing.

References

Ackerman, S. et al. (1992). *Discovering the brain*. National Academies Press. Cited on page 6.

Adams, P., Rabbi, M., Rahman, T., Matthews, M., Voida, A., Gay, G., Choudhury, T., and Voida, S. (2014). Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild. *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. Cited on pages 12 and 13.

Agrafioti, F., Hatzinakos, D., Member, S., and Anderson, A. K. (2012). ECG Pattern Analysis for Emotion Detection.pdf. 3(1):102–115. Cited on pages 2 and 13.

Al-Nafjan, A., Hosny, M., Al-Ohali, Y., and Al-Wabil, A. (2017). Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review. *Applied Sciences*, 7(12):1239. Cited on pages 3, 16, 19, and 24.

Al-Shargie, F., Kiguchi, M., Badruddin, N., Dass, S. C., Hani, A. F. M., and Tang, T. B. (2016). Mental stress assessment using simultaneous measurement of EEG and fNIRS. *Biomedical Optics Express*, 7(10):3882. Cited on page 17.

Al'Absi, M. and Arnett, D. (2000). Adrenocortical responses to psychological stress and risk for hypertension. *Biomedicine & pharmacotherapy*, 54(5):234–244. Cited on page 2.

Alarcao, S. M. and Fonseca, M. J. (2017). Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*, 3045(c):1–20. Cited on pages 16, 19, 21, 22, 24, 25, 31, 36, and 40.

Alberdi, A., Aztiria, A., and Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, 59:49–75. Cited on pages 11, 12, 14, and 17.

Allwein, E. L., Schapire, R. E., and Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of machine learning research*, 1(Dec):113–141. Cited on page 35.

Alotaiby, T., El-Samie, F. E. A., Alshebeili, S. A., and Ahmad, I. (2015). A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):66. Cited on page 7.

American Psychological Association (2015). The State-Trait Anxiety Inventory (STAI). [Online] `http://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/trait-state.aspx` (Accessed: 30-12-2017). Cited on page 11.

Arnrich, B., Setz, C., La Marca, R., Tröster, G., and Ehlert, U. (2010). What does your chair know about your stress level? *IEEE Transactions on Information Technology in Biomedicine*, 14(2):207–214. Cited on page 12.

Athavale, Y. and Krishnan, S. (2017). Biosignal monitoring using wearables: Observations and opportunities. *Biomedical Signal Processing and Control*, 38:22–33. Cited on page 27.

Atkinson, J. and Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications*, 47:35–41. Cited on page 1.

Bansal, D., Khan, M., and Salhan, A. (2009). A review of measurement and analysis of heart rate variability. In *Computer and Automation Engineering, 2009. ICCAE'09. International Conference on*, pages 243–246. IEEE. Cited on page 14.

Bao, A.-M., Meynen, G., and Swaab, D. (2008). The stress system in depression and neurodegeneration: focus on the human hypothalamus. *Brain research reviews*, 57(2):531–553. Cited on page 5.

Baumgartner, T., Valko, L., Esslen, M., and Jäncke, L. (2006). Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an eeg and psychophysiology study. *Cyber-Psychology & Behavior*, 9(1):30–45. Cited on pages 28 and 40.

Bhide, A., Durgaprasad, R., Kasala, L., Velam, V., and Hulikal, N. (2016). Electrocardiographic changes during acute mental stress. *International Journal of Medical Science and Public Health*, 5(5):835–838. Cited on page 15.

Bradley, M. and Lang, P. J. (1994). Measuring Emotion : The Self-Assessment Manikin and the Semantic Differential. 25(I). Cited on pages 10 and 11.

Bradley, M. M. and Lang, P. J. (2007). The international affective digitized sounds (; iads-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3*. Cited on page 9.

Cacioppo, J., Tassinary, L. G., and Berntson, G. G. (2007). *The Handbook of Psychophysiology*, volume 44. Cambridge University Press, 3rd editio edition. Cited on pages 14 and 18.

Calibo, T. K., Blanco, J. A., and Firebaugh, S. L. (2013). Cognitive stress recognition. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pages 1471–1475. Cited on pages 3, 4, and 10.

Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., and Su, S. (2015). Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015-Novem(November):7250–7253. Cited on pages 32 and 38.

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, 26(3):321–352. Cited on page 7.

Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., and Pecchia, L. (2015). Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control*, 18:370–377. Cited on page 13.

Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology*, 5(7):374–381. Cited on page 7.

Chrousos, G. P. and Gold, P. W. (1992). The Concepts of Stress and Stress System Disorders: Overview of Physical and Behavioral Homeostasis. *JAMA: The Journal of the American Medical Association*, 267(9):1244–1252. Cited on page 2.

Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A Global Measure of Perceived Stress. Cited on page 10.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297. Cited on page 25.

Coyle, S. M., Ward, T. E., and Markham, C. M. (2007). Brain-computer interface using a simplified functional near-infrared spectroscopy system. *Journal of neural engineering*, 4(3):219–226. Cited on page 17.

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2):132–142. Cited on page 17.

Critchley, H. D. (2009). Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *International Journal of Psychophysiology*, 73(2):88–94. Cited on page 1.

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, 5(7):583. Cited on pages 6 and 7.

Davidson, R. J. (2004). What does the prefrontal cortex "do" in affect: perspectives on frontal eeg asymmetry research. *Biological psychology*, 67(1-2):219–234. Cited on page 40.

De Kloet, E. R., Joëls, M., and Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience*, 6(6):463–475. Cited on page 43.

DEAP (2012). DEAP: A Dataset for Emotion Analysis using Physiological and Audiovisual Signals. [Online] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/readme.html (Accessed: 02-05-2018). Cited on page 32.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21. Cited on page 52.

Dhaka, S. and Kashyap, N. (2017). Explicit emotion regulation: Comparing emotion inducing stimuli. *Psychological Thought*, 10(2):303–314. Cited on pages 28 and 55.

Dietterich, T. G. and Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, 2:263–286. Cited on page 35.

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., Goldenstein, S. K., Vogler, C., Venkataraman, S., and Metaxas, D. N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine*, 76(6):B172–B182. Cited on pages 12 and 13.

Djamshidian, A. and Lees, A. J. (2014). Can stress trigger Parkinson's disease? *Journal of Neurology, Neurosurgery and Psychiatry*, 85(8):879–882. Cited on page 2.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., and Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4):712–717. Cited on page 1.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2):179–211. Cited on page 26.

EMOTIV (2018). EMOTIV EPOC+ - 14 Channel Wireless EEG Headset. [Online] https://www.emotiv.com/epoc/ (Accessed: 28-01-2018). Cited on page 51.

Epp, C., Lippold, M., and Mandryk, R. L. (2011). Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 715–724. ACM. Cited on page 12.

Everly, G. S. and Lating, J. M. (2013). The Anatomy and Physiology of the Human Stress Response. In *A Clinical Guide to the Treatment of the Human Stress Response*, volume 254, pages 17–51. Springer New York, New York, NY. Cited on pages 5 and 13.

Fazli, S., Mehnert, J., Steinbrink, J., and Blankertz, B. (2012). Using NIRS as a predictor for eeg-based bci performance. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4911–4914. IEEE. Cited on page 17.

Fernandez, R. and Picard, R. W. (2003). Modeling drivers' speech under stress. *Speech communication*, 40(1):145–159. Cited on page 12.

Furness, J. B. (2006). The organisation of the autonomic nervous system: peripheral connections. *Autonomic Neuroscience: Basic and Clinical*, 130(1):1–5. Cited on page 5.

Gabert-Quillen, C. A., Bartolini, E. E., Abravanel, B. T., and Sanislow, C. A. (2015). Ratings for emotion film clips. *Behavior Research Methods*, 47(3):773–787. Cited on pages 9 and 20.

Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., Crifaci, G., Vetrano, N., Giulintano, A., Bernava, G., et al. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. *Journal of medical Internet research*, 16(7). Cited on page 3.

Galán, F., Nuttin, M., Lew, E., Ferrez, P. W., Vanacker, G., Philips, J., and Millán, J. d. R. (2008). A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. *Clinical neurophysiology*, 119(9):2159–2169. Cited on page 3.

Gartner (2017). Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017. Cited on page 3.

Haapalainen, E., Kim, S., Forlizzi, J. F., and Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10*, page 301. Cited on page 18.

Hagmueller, M., Rank, E., and Kubin, G. (2006). Evaluation of the Human Voice for Indications of Workload Induced Stress in the Aviation Environment. (18):89. Cited on page 12.

Hall, J. E. (2015). *Guyton and Hall Textbook of Medical Physiology E-Book*. Elsevier Health Sciences. Cited on pages 7 and 16.

Ham, J., Cho, D., Oh, J., and Lee, B. (2017). Discrimination of Multiple Stress Levels in Virtual Reality Environments Using Heart Rate Variability. pages 3989–3992. Cited on pages 13 and 14.

Hassard, J., Teoh, K., Dewe, P., Cosmar, M., Grundler, R., Flemming, D., Cosemans, B., and Van den Broek, K. (2014). Calculating the cost of work-related stress and psychosocial risks. Technical report, European Agency for Safety and Health at Work. Cited on page 2.

Hayashi, T., Mizuno-Matsumoto, Y., Okamoto, E., Kato, M., and Murata, T. (2012). An fMRI study of brain processing related to stress states. In *World Automation Congress (WAC), 2012*, pages 1–6. IEEE. Cited on page 17.

Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6(2):156–166. Cited on pages 10, 14, 17, and 18.

Hellhammer, D. H., Wüst, S., and Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, 34(2):163–171. Cited on page 13.

Hernandez, J., Paredes, P., Roseway, A., and Czerwinski, M. (2014). Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–60. ACM. Cited on page 17.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283. Cited on pages 23 and 34.

Hines, E. A. and Brown, G. E. (1936). The cold pressor test for measuring the reactibility of the blood pressure: data concerning 571 normal and hypertensive subjects. *American heart journal*, 11(1):1–9. Cited on page 9.

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., and Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, 92(1-2):84–89. Cited on page 15.

Horlings, R., Datcu, D., and Rothkrantz, L. (2008). Emotion recognition using brain activity. *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, (March):6. Cited on page 40.

Hosseini, S. and Khalilzadeh, M. (2010). Emotional stress recognition system using EEG and psychophysiological signals:. *Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on*, pages 1 – 6. Cited on pages 4, 16, 20, 21, 22, 23, 24, 26, and 32.

Hou, X., Liu, Y., Sourina, O., Tan, Y. R. E., Wang, L., and Mueller-Wittig, W. (2015). EEG Based Stress Monitoring. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, (November):3110–3115. Cited on pages 10, 20, and 23.

Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375. Cited on page 16.

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, 5(3):327–339. Cited on pages 22, 24, 34, 35, and 40.

Jie, X., Cao, R., and Li, L. (2014). Emotion recognition based on the sample entropy of EEG. *Bio-Medical Materials and Engineering*, 24(1):1185–1192. Cited on pages 44 and 45.

Jun, G. and Smitha, K. G. (2016). EEG based Stress Level Identification. *IEEE International Conference on Systems, Man, and Cybernetics*, pages 3270–3274. Cited on pages 4, 10, 20, and 23.

Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M., and Madsen, J. R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature*, 399(6738):781. Cited on page 56.

Kaklauskas, A., Zavadskas, E., Seniut, M., Dzemyda, G., Stankevic, V., Simkevičius, C., Stankevic, T., Paliskiene, R., Matuliauskaite, A., Kildiene, S., Bartkiene, L., Ivanikovas, S., and Gribniak, V. (2011). Web-based Biometric Computer Mouse Advisory System to Analyze a User's Emotions and Work Productivity. *Engineering Applications of Artificial Intelligence*, 24(6):928–945. Cited on page 15.

Katsigiannis, S. and Ramzan, N. (2018). DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1):98–107. Cited on page 20.

Khansari, D. N., Murgo, A. J., and Faith, R. E. (1990). Effects of stress on the immune system. *Immunology today*, 11:170–175. Cited on page 2.

Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short term monitoring of physiological signals. *Medical Biological Engineering and computing*, 42(Journal Article):419–427. Cited on page 1.

Kirschbaum, C., Pirke, K.-M., and Hellhammer, D. H. (1993). The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81. Cited on page 9.

Kober, S. E., Kurzmann, J., and Neuper, C. (2012). Cortical correlate of spatial presence in 2D and 3D interactive virtual reality : An EEG study. *International Journal of Psychophysiology*, 83(3):365–374. Cited on pages 28, 40, and 56.

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). DEAP: A dataset for emotion analysis using physiological and audiovisual signals. *IEEE Transactions on Affective Computing*, 3:1–15. Cited on pages 1, 19, 22, and 32.

Koh, K. B., Park, J. K., Kim, C. H., and Cho, S. (2001). Development of the stress response inventory and its application in clinical practice. *Psychosomatic medicine*, 63(4):668–78. Cited on page 10.

Kolakowska, A. (2013). A review of emotion recognition methods based on keystroke dynamics and mouse movements. *2013 6th International Conference on Human System Interactions, HSI 2013*, pages 548–555. Cited on page 12.

Kragel, P. A. and LaBar, K. S. (2016). Decoding the Nature of Emotion in the Brain. *Trends in Cognitive Sciences*, 20(6):444–455. Cited on page 7.

Kumar, D., Arjunan, S. P., and Aliahmad, B. (2017). *Fractals: applications in biological signalling and image processing*. CRC Press. Cited on page 23.

Kweon, S. H., Kweon, H. J., Kim, S.-j., Li, X., Liu, X., and Kweon, H. L. (2017). A brain wave research on vr (virtual reality) usage: Comparison between vr and 2d video in eeg measurement. In *International Conference on Applied Human Factors and Ergonomics*, pages 194–203. Springer. Cited on page 28.

Lang, P., Bradley, M., and Cuthbert, B. (2008). International Affective Picture System (IAPS): Technical Manual and Affective Ratings. Technical report, University of Florida, Gainesville, FL. Cited on pages 9, 48, and 49.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, 50(5):372. Cited on page 1.

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual review of psychology*, 44(1):1–22. Cited on page 1.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual review of neuroscience*, 23(1):155–184. Cited on page 6.

Lee, Y. Y. and Hsieh, S. (2014). Classifying different emotional states by means of eegbased functional connectivity patterns. *PLoS ONE*, 9(4). Cited on pages 24 and 40.

Lentz, T., Ratcliff, G., Loewy, A., Nathan, P., Rudge, P., Noback, C., Matthews, P., and Haines, D. (2017). Encyclopædia Britannica: Human nervous system. [Online] `https://www.britannica.com/science/human-nervous-system` (Accessed: 30-12-2017). Cited on page 5.

Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., and Andreoli, A. (1993). Development of the perceived stress questionnaire: A new tool for psychosomatic research. Cited on page 10.

Li, B. J., Bailenson, J. N., Pines, A., Greenleaf, W. J., and Williams, L. M. (2017). A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Frontiers in Psychology*, 8(DEC). Cited on pages 50 and 52.

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018). Exploring EEG features in cross-subject emotion recognition. *Frontiers in Neuroscience*, 12(MAR). Cited on pages 22, 26, 39, 40, 44, and 45.

Lin, Y.-P., Yang, Y.-H., and Jung, T.-P. (2014). Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening. *Frontiers in Neuroscience*, 8(8 MAY):1–14. Cited on pages 44 and 45.

Liu, Y. and Sourina, O. (2013). EEG databases for emotion recognition. *Proceedings - 2013 International Conference on Cyberworlds, CW 2013*, (October 2013):302–309. Cited on page 40.

Liu, Y. and Sourina, O. (2016). Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm Real-time Subject-dependent EEG-based Emotion Recognition Algorithm. (August). Cited on pages 21, 26, and 45.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., and Choudhury, T. (2012). StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, page 351. Cited on pages 12 and 13.

Lupien, S. J. and Seguin, F. (2013). How to measure stress in humans. *Centre for Studies in Human Stress*. Cited on page 13.

MacLean, P. D. (1955). The limbic system (visceral brain) and emotional behavior. *AMA Archives of Neurology & Psychiatry*, 73(2):130–134. Cited on page 6.

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2):314–324. Cited on page 48.

Matlovič, T. (2016). Emotion Detection using EPOC EEG device. pages 1–6. Cited on pages 16, 20, 22, 24, and 26.

McClellan, J., Parks, T., and Rabiner, L. (1973). A computer program for designing optimum fir linear phase digital filters. *IEEE Transactions on Audio and Electroacoustics*, 21(6):506–526. Cited on page 32.

McDuff, D., Karlson, A., Kapoor, A., Roseway, A., and Czerwinski, M. (2012). Affectaura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 849–858. ACM. Cited on page 12.

McEwen, B. S. and Stellar, E. (1993). Stress and the individual: mechanisms leading to disease. *Archives of internal medicine*, 153(18):2093–2101. Cited on page 2.

Mehmood, R. M. and Lee, H. J. (2016). A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns. *Computers and Electrical Engineering*, 53:444–457. Cited on pages 20, 21, 26, and 45.

Menezes, M. L., Samara, A., Galway, L., Sant'Anna, A., Verikas, A., Alonso-Fernandez, F., Wang, H., and Bond, R. (2017). Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset. *Personal and Ubiquitous Computing*, pages 1–11. Cited on pages 20, 22, 23, 24, 26, and 34.

Milczarek, M. (2014). Managing stress and psychosocial risks at work. Technical Report October, European Agency for Safety and Health at Work. Cited on page 2.

Millán, J. d. R., Rupp, R., Mueller-Putz, G., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kubler, A., Leeb, R., et al. (2010). Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in neuroscience*, 4:161. Cited on page 3.

Müller, K. R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1):82–90. Cited on page 15.

Nakisa, B., Rastgoo, M. N., Tjondronegoro, D., and Chandran, V. (2018). Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Systems with Applications*, 93:143–155. Cited on pages 22, 23, 24, 27, 34, and 45.

Nicolaides, N. C., Kyratzi, E., Lamprokostopoulou, A., Chrousos, G. P., and Charmandari, E. (2015). Stress, the Stress System and the Role of Glucocorticoids. *Neuroimmunomodulation*, 22(1-2):6–19. Cited on pages 2 and 14.

Paiva, J. S., Rodrigues, S., and Cunha, J. P. S. (2016). Changes in ST, QT and RR ECG intervals during acute stress in firefighters: A pilot study. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016-Octob:3378–3381. Cited on page 15.

Parisi, T. (2015). *Learning virtual reality: developing immersive experiences and applications for desktop, web, and mobile*. " O'Reilly Media, Inc.". Cited on page 27.

Patton, K. T. (2015). *Anatomy and Physiology-E-Book*. Elsevier Health Sciences. Cited on page 5.

Pessoa, L. (2018). Understanding emotion with brain networks. *Current Opinion in Behavioral Sciences*, 19:19–25. Cited on page 7.

Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):186–197. Cited on pages 23 and 34.

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., and Strohecker, C. (2004). Affective learning—a manifesto. *BT technology journal*, 22(4):253–269. Cited on page 2.

Pomer-Escher, A. G., de Souza, M. D. P., and Filho, T. F. B. (2014). Methodology for analysis of stress level based on asymmetry pattersn of alpha rhythms in EEG signals. *Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living, 5th ISSNIP-IEEE*, pages 1–5. Cited on page 4.

Rahman, M. M., Epstein, D. H., Preston, K. L., Jobes, M., Beck, J. G., Kedia, S., Ward, K. D., Al'Absi, M., Kumar, S., Bari, R., Ali, A. A., Sharmin, M., Raij, A., Hovsepian, K., Hossain, S. M., Ertin, E., and Kennedy, A. (2014). Are we there yet? Feasibility of continuous stress assessment via wireless physiological sensors. *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14*, 2014:479–488. Cited on pages 3 and 13.

Rathod, P., George, K., and Shinde, N. (2016). Bio-signal based emotion detection device. *BSN 2016 - 13th Annual Body Sensor Networks Conference*, pages 105–108. Cited on page 12.

Rey, B. and Alcañiz, M. (2010). Research in neuroscience and virtual reality. In *Virtual Reality*. InTech. Cited on page 27.

Robnik-Šikonja, M. and Kononenko, I. (2003). Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*, 53(1-2):23–69. Cited on page 35.

Rothman, S. M. and Mattson, M. P. (2010). Adverse stress, hippocampal networks, and Alzheimer's disease. *Neuromolecular medicine*, 12(1):56–70. Cited on page 2.

Rowland, N., Meile, M., Nicolaidis, S., et al. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, 228(4700):750–752. Cited on page 16.

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social science information*, 44(4):695–729. Cited on page 1.

Selvaraj, J., Murugappan, M., Wan, K., and Yaacob, S. (2013). Classification of emotional states from electrocardiogram signals: a non-linear approach based on hurst. *BioMedical Engineering OnLine*, 12(1):44. Cited on pages 2, 3, and 13.

Selye, H. (1956). *The stress of life*. Cited on page 1.

Seo, S.-H. and Lee, J.-T. (2010). Stress and EEG. In *Convergence and Hybrid Information Technologies*. InTech. Cited on pages 13 and 16.

Sharma, N. and Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, 108(3):1287–1301. Cited on pages 4, 11, 12, 14, and 18.

Shin, D., Shin, D., and Shin, D. (2017). Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents. *Multimedia Tools and Applications*, 76(9):11449–11470. Cited on pages 23 and 26.

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., Thoma, M. V., Spoerri, C., Ehlert, U., and Nater, U. M. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology*, 51:227–236. Cited on page 10.

Slobounov, S. M., Ray, W., Johnson, B., Slobounov, E., and Newell, K. M. (2015). Modulation of cortical activity in 2D versus 3D virtual reality environments : An EEG study. *International Journal of Psychophysiology*, 95(3):254–260. Cited on pages 28 and 56.

Sohaib, A. T., Qureshi, S., Hagelbäck, J., Hilborn, O., and Jerčić, P. (2013). Evaluating classifiers for emotion recognition using eeg. In *International Conference on Augmented Cognition*, pages 492–501. Springer. Cited on pages 44 and 45.

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55. Cited on page 20.

Sourina, O. and Liu, Y. (2011). A Fractal-Based Algorithm of Emotion Recognition from EEG using Arousal-Valence Model. *Proc of Biosignals*, pages 209–214. Cited on page 34.

Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). Manual for the state-trait anxiety inventory. Cited on page 11.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643. Cited on page 9.

Subhani, A. R., Mumtaz, W., Saad, M. N. B. M., Kamel, N., and Malik, A. S. (2017). Machine Learning Framework for the Detection of Mental Stress at Multiple Levels. *IEEE Access*, 5:13545–13556. Cited on pages 16, 20, and 23.

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. *Monographs of the society for research in child development*, 59(2-3):25–52. Cited on page 1.

Tomita, Y., Vialatte, F. B., Dreyfus, G., Mitsukura, Y., Bakardjian, H., and Cichocki, A. (2014). Bimodal BCI using simultaneously NIRS and EEG. *IEEE Transactions on Biomedical Engineering*, 61(4):1274–1284. Cited on page 17.

Uhrig, M. K., Trautmann, N., Baumgärtner, U., Treede, R.-D., Henrich, F., Hiller, W., and Marschall, S. (2016). Emotion elicitation: A comparison of pictures and films. *Frontiers in psychology*, 7:180. Cited on pages 28 and 55.

Urbanowicz, R. J., Meeker, M., LaCava, W., Olson, R. S., and Moore, J. H. (2017). Relief-based feature selection: introduction and review. *arXiv preprint:1711.08421*. Cited on pages 35 and 43.

Valenzi, S., Islam, T., Jurica, P., and Cichocki, A. (2014). Individual Classification of Emotions Using EEG. *Journal of Biomedical Science and Engineering*, 07(08):604–620. Cited on pages 9, 16, 20, 22, 24, 26, 32, and 45.

Van Boxtel, A. (2010). Facial EMG as a tool for inferring affective states. In *Proceedings of measuring behavior*, pages 104–108. Noldus Information Technology Wageningen. Cited on page 18.

Wang, X. W., Nie, D., and Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129:94–106. Cited on pages 32 and 40.

Wei, C. Z. (2013). Stress emotion recognition based on RSP and EMG signals. In *Advanced Materials Research*, volume 709, pages 827–831. Trans Tech Publ. Cited on page 10.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73. Cited on page 33.

Wijsman, J., Grundlehner, B., Liu, H., Penders, J., and Hermens, H. (2013). Wearable physiological sensors reflect mental stress state in office-like situations. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 600–605. IEEE. Cited on pages 10 and 18.

Wood, D. P., Murphy, J., Center, K., McLay, R., Reeves, D., Pyne, J., Shilling, R., and Wiederhold, B. K. (2006). Combat-related post-traumatic stress disorder: A case report using virtual reality exposure therapy with physiological monitoring. *Cyberpsychology & Behavior*, 10(2):309–315. Cited on page 3.

Zhang, J., Chen, M., Zhao, S., Hu, S., Shi, Z., and Cao, Y. (2016). ReliefF-based EEG sensor selection methods for emotion recognition. *Sensors (Switzerland)*, 16(10):1–15. Cited on pages 24, 35, 39, and 40.

Zheng, W.-L., Zhu, J.-Y., and Lu, B.-L. (2017). Identifying stable patterns over time for emotion recognition from eeg. *IEEE Transactions on Affective Computing*. Cited on page 20.